# GUIDANCE ON MONITORING OF MERCURY AND MERCURY COMPOUNDS TO SUPPORT THE EFFECTIVENESS EVALUATION OF THE MINAMATA CONVENTION

*Draft of 15 July 2021*

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF ACRONYMS</td>
<td>5</td>
</tr>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION AND OBJECTIVES</td>
<td>16</td>
</tr>
<tr>
<td>1.1. INTRODUCTION</td>
<td>16</td>
</tr>
<tr>
<td>1.2. OBJECTIVES</td>
<td>16</td>
</tr>
<tr>
<td>CHAPTER 2. COMPARABLE MONITORING DATA AND THE EFFECTIVENESS EVALUATION</td>
<td>18</td>
</tr>
<tr>
<td>2.1. INTRODUCTION</td>
<td>18</td>
</tr>
<tr>
<td>2.2. WEIGHT OF EVIDENCE AND OPERATIONAL QUESTIONS</td>
<td>19</td>
</tr>
<tr>
<td>2.3. MONITORING MATRICES</td>
<td>21</td>
</tr>
<tr>
<td>2.4. TIERED APPROACH FOR DEVELOPING AND IMPROVING MONITORING PROGRAMMES</td>
<td>23</td>
</tr>
<tr>
<td>2.5. QUALITY OF MONITORING DATA</td>
<td>24</td>
</tr>
<tr>
<td>2.6. DATA MANAGEMENT</td>
<td>25</td>
</tr>
<tr>
<td>CHAPTER 3. ATMOSPHERIC MERCURY MONITORING</td>
<td>27</td>
</tr>
<tr>
<td>3.1. INTRODUCTION</td>
<td>27</td>
</tr>
<tr>
<td>3.2. SIGNIFICANCE OF AIR AS A MATRIX FOR MERCURY MONITORING</td>
<td>27</td>
</tr>
<tr>
<td>3.2.1. Mercury in air</td>
<td>28</td>
</tr>
<tr>
<td>3.2.1. Atmospheric mercury deposition</td>
<td>29</td>
</tr>
<tr>
<td>3.3. TIERED APPROACH FOR ATMOSPHERIC MERCURY MONITORING</td>
<td>29</td>
</tr>
<tr>
<td>3.4. WHERE SHOULD ATMOSPHERIC MERCURY MEASUREMENTS BE COLLECTED</td>
<td>30</td>
</tr>
<tr>
<td>3.5. HOW TO MEASURE MERCURY IN AIR: SAMPLING AND MEASUREMENT METHODS</td>
<td>31</td>
</tr>
<tr>
<td>3.5.1. Active air sampling</td>
<td>32</td>
</tr>
<tr>
<td>3.5.2. Passive sampling</td>
<td>33</td>
</tr>
<tr>
<td>3.5.3. Wet deposition sampling</td>
<td>34</td>
</tr>
<tr>
<td>3.5.4. Dry deposition sampling</td>
<td>35</td>
</tr>
<tr>
<td>3.6. FREQUENCY AND DURATION OF SAMPLING</td>
<td>35</td>
</tr>
<tr>
<td>3.7. QUALITY ASSURANCE AND QUALITY CONTROL FOR FIELD AIR MONITORING OPERATIONS</td>
<td>36</td>
</tr>
<tr>
<td>3.8. MANAGEMENT, ANALYSIS AND EVALUATION OF ATMOSPHERIC MERCURY DATA</td>
<td>37</td>
</tr>
<tr>
<td>3.8.1. Local, regional and hemispheric trend analysis</td>
<td>37</td>
</tr>
<tr>
<td>3.8.2. Data based analysis</td>
<td>37</td>
</tr>
<tr>
<td>3.8.3. Source receptor relationship based on footprints and trajectory analysis</td>
<td>38</td>
</tr>
<tr>
<td>3.9. CONCLUSIONS</td>
<td>38</td>
</tr>
</tbody>
</table>
The Secretariat of the Minamata Convention wishes to acknowledge and thank the countries, institutions and experts who made the development of this guidance possible. First among those are over 80 experts – scientists from chemical, biological and human sciences, policymakers, Indigenous representatives, and health practitioners, amongst others – who have given freely their time and expertise to shape and develop this document, acting as peer-reviewers of the successive drafts of this document.

A special thanks go to the four Lead Authors of the different chapters – Lynwill Martin (chapter 3), David Evers (chapter 4), Niladri Basu (chapter 5) and Colin Thackray (chapter 6) – who have each gone well beyond the call of duty and donated a substantial contribution of time, effort and intellectual capital to bring each chapter to fruition, over the last two years and many drafts and short deadlines.

We express our gratitude for the support provided by other UN organizations, including the World Health Organization (WHO), the International Labour Organization (ILO), the International Atomic Energy Agency (IAEA), UN Environment Programme (UNEP) and the Secretariat of the Basel, Rotterdam and Stockholm Conventions, as well as to various national Governments that have participated in the different rounds of review and helped improve the guidance with a multidisciplinary focus.

Last, but not least, appreciation goes to all members of the Minamata Convention family who were actively engaged throughout the process to conceptualize and develop this guidance.
List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAS</td>
<td>Atomic absorption spectrometry</td>
</tr>
<tr>
<td>AFS</td>
<td>Atomic fluorescence spectrometry</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>AMAP</td>
<td>Arctic Monitoring and Assessment Programme</td>
</tr>
<tr>
<td>ASGM</td>
<td>Artisanal and small-scale gold mining</td>
</tr>
<tr>
<td>CARE</td>
<td>Collective Benefits, Authority to Control, Responsibility, and Ethics</td>
</tr>
<tr>
<td>CART</td>
<td>Classification and regression tree</td>
</tr>
<tr>
<td>CHMS</td>
<td>Canadian Health Measures Survey</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CIOMS</td>
<td>Council for International Organizations of Medical Sciences</td>
</tr>
<tr>
<td>COP</td>
<td>Conference of the Parties</td>
</tr>
<tr>
<td>CV</td>
<td>Cold vapour</td>
</tr>
<tr>
<td>DHS</td>
<td>Demographic and Health Surveys</td>
</tr>
<tr>
<td>DMA</td>
<td>Direct mercury analyzer</td>
</tr>
<tr>
<td>DOC</td>
<td>Dissolved Organic Carbon</td>
</tr>
<tr>
<td>dw</td>
<td>Dry weight</td>
</tr>
<tr>
<td>EHMS</td>
<td>Environmental Health Monitoring System</td>
</tr>
<tr>
<td>FAIR</td>
<td>Findability, Accessibility, Interoperability, and Reuse</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization</td>
</tr>
<tr>
<td>fw</td>
<td>Fresh weight</td>
</tr>
<tr>
<td>fww</td>
<td>Fresh wet weight</td>
</tr>
<tr>
<td>GAM</td>
<td>Generalized additive model</td>
</tr>
<tr>
<td>GEF</td>
<td>Global Environment Facility</td>
</tr>
<tr>
<td>GEM</td>
<td>Gaseous elemental mercury</td>
</tr>
<tr>
<td>GerES</td>
<td>German Environmental Survey</td>
</tr>
<tr>
<td>GLM</td>
<td>General linear model</td>
</tr>
<tr>
<td>GLMM</td>
<td>Generalized linear mixed model</td>
</tr>
<tr>
<td>GOM</td>
<td>Gaseous oxidised mercury</td>
</tr>
<tr>
<td>HBM</td>
<td>Human biomonitoring</td>
</tr>
<tr>
<td>Hg</td>
<td>Mercury</td>
</tr>
<tr>
<td>IAEA</td>
<td>International Atomic Energy Agency</td>
</tr>
<tr>
<td>ILO</td>
<td>International Labour Organization</td>
</tr>
<tr>
<td>INSPQ</td>
<td>Institut national de santé publique du Québec</td>
</tr>
<tr>
<td>IUCN</td>
<td>International Union for Conservation of Nature</td>
</tr>
<tr>
<td>KoNEHS</td>
<td>Korean National Environmental Health Survey</td>
</tr>
<tr>
<td>LSM</td>
<td>Large-scale mining</td>
</tr>
<tr>
<td>MeHg</td>
<td>Methylmercury</td>
</tr>
<tr>
<td>MK</td>
<td>Mann-Kendall</td>
</tr>
<tr>
<td>MRPP</td>
<td>Multiple-response permutation procedure</td>
</tr>
<tr>
<td>NCP</td>
<td>Northern Contaminants Program</td>
</tr>
<tr>
<td>NHANES</td>
<td>National Health and Nutrition Examination Survey</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>NIES</td>
<td>National Institute for Environmental Studies, Japan</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NTP</td>
<td>National Toxicology Program</td>
</tr>
<tr>
<td>OCAP</td>
<td>Ownership, Control, Access, and Possession</td>
</tr>
<tr>
<td>PAS</td>
<td>Passive air sampler</td>
</tr>
<tr>
<td>PBM</td>
<td>Particle-bound mercury</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PMF</td>
<td>Probability mass function</td>
</tr>
<tr>
<td>PSCF</td>
<td>Potential source contribution function</td>
</tr>
<tr>
<td>QA/QC</td>
<td>Quality assurance and quality control</td>
</tr>
<tr>
<td>RGM</td>
<td>Reactive gaseous mercury</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SOC</td>
<td>Soil organic carbon</td>
</tr>
<tr>
<td>SOP</td>
<td>Standard operating procedure</td>
</tr>
<tr>
<td>SSR</td>
<td>Sample submission report</td>
</tr>
<tr>
<td>STROBE</td>
<td>Strengthening the reporting of observational studies</td>
</tr>
<tr>
<td>TGM</td>
<td>Total gaseous mercury</td>
</tr>
<tr>
<td>THg</td>
<td>Total mercury</td>
</tr>
<tr>
<td>TSS</td>
<td>Total suspended solids</td>
</tr>
<tr>
<td>UNDP</td>
<td>United Nations Development Organization</td>
</tr>
<tr>
<td>UNEP</td>
<td>United Nations Environment Programme</td>
</tr>
<tr>
<td>UNIDO</td>
<td>United Nations Industrial Development Organization</td>
</tr>
<tr>
<td>US CDC</td>
<td>United States Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>US EPA</td>
<td>United States Environmental Protection Agency</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WMO</td>
<td>World Meteorological Organization</td>
</tr>
<tr>
<td>ww</td>
<td>Wet weight</td>
</tr>
</tbody>
</table>
Executive Summary

Paragraph 2 of Article 22 on Effectiveness Evaluation of the Minamata Convention requires the Conference of the Parties (COP) to make “arrangements for providing itself with comparable monitoring data on the presence and movement of mercury and mercury compounds in the environment as well as trends in levels of mercury and mercury compounds observed in biotic media and vulnerable populations”.

This document provides scientific and technical guidance to support the COP to obtain comparable monitoring data for the Effectiveness Evaluation. The primary objectives of this document are to (i) explain the role of monitoring in the Effectiveness Evaluation and set realistic expectations about what can be learned over time, (ii) provide guidance to Parties and organizations, which are currently conducting monitoring programmes, on what data and accompanying information would inform the Effectiveness Evaluation, and (iii) provide guidance to Parties and organizations who wish to develop new monitoring programs, or improve existing ones, with a view to contributing to the Effectiveness Evaluation.

Building on four overarching policy questions,

- Have the Parties taken actions to implement the Minamata Convention?
- Have the actions taken resulted in changes in mercury supply, use, emissions and releases into the environment?
- Have those changes resulted in changes in levels of mercury in the environment, biotic media and vulnerable populations that can be attributed to the Minamata Convention?
- To what extent are existing measures under the Minamata Convention meeting the objective of protecting human health and the environment from mercury?

This document describes the scientific and technical processes and guiding principles for compiling and/or generating comparable monitoring data, as well as methods to use such monitoring data for understanding the presence, movements and trends of mercury in the environment and humans. Throughout the guidance, monitoring activities to support the Effectiveness Evaluation have been grouped into five categories:

1) Estimation of contemporary mercury concentrations for areas without (i.e. background sites) or with (i.e. affected sites) local anthropogenic sources;
2) Identification of temporal trends;
3) Characterization of spatial patterns;
4) Estimation of source attribution;
5) Estimation of exposure and adverse impacts;
6) Quantification of key environmental processes to improve our understanding of cause-effect relationships.

From each of these categories, operational questions can be drawn to guide the collection and analysis of the relevant monitoring data and inform the Effectiveness Evaluation in complementary ways. The continuum of scientific weight of evidence that the operational questions provide is reflected in the
three tiers presented in this guidance. Together, answers to the operational questions form a successive continuum of scientific weight of evidence that can form an evidentiary basis for supporting the Effectiveness Evaluation. The operational questions are as follows.

1) Estimation of contemporary mercury concentrations for areas without (i.e. background sites) or with (i.e. affected sites) local anthropogenic sources
   - What are the levels and form of mercury in sites that are expected to be remote from anthropogenic sources?
   - What are the levels and form of mercury in sites that are expected to be affected by local anthropogenic point sources?

2) Identification of temporal trends
   - Do the levels and form of mercury in the observed matrix (air, biota, human) at a given location change over time – for example, in short- (< 5 years), medium- (5 to 20 years), and long-term periods (> 20 years)? Is there a long-term trend or trajectory (a signal) that can be separated from the temporal variability (noise)?
   - How do observed temporal variations and trends differ spatially?
   - How do observed temporal variations and trends in mercury compare to or co-vary with variations and trends of:
     - Mercury in different forms (chemical species) or other within matrices?
     - Mercury emissions and releases?
     - Related pollutants/emissions or environmental variables?

3) Characterization of spatial patterns
   - What are the levels and form of mercury in the observed matrix (air, biota, human) at a given location and time?
   - Taken together, what does the available data suggest about:
     - Spatial variability in environmental mercury concentrations?
     - Variability in environmental mercury concentrations within populations, their habitats, and ecosystems?
   - Do the observed spatial variations and patterns or gradients differ among:
     - Forms (chemical species) of mercury?
     - Environmental matrices?
   - How do the observed spatial variations and patterns or gradients compare to those of:
     - Mercury emissions and releases?
     - Related pollutants/emissions or environmental variables?

4) Estimation of source attribution
   - Using models and statistical analyses consistent with observational data, how can the observed levels, spatial patterns, temporal trends and adverse impacts on species, ecosystem services, biodiversity and human populations be attributed to changes:
     - In sources or sinks of anthropogenic, natural and legacy mercury?
In anthropogenic sources (local, regional, global) of mercury?

- Influenced by the Convention?
  - Not influenced by the Convention?

5) Estimation of exposure and adverse impacts

- How do mercury concentrations impact populations, species, ecosystem services, biodiversity and human populations in regions that are remote from sources as well as those that are locally impacted by anthropogenic sources?
- How do the observed levels of mercury in air, biota, and humans compare to established national and international benchmark levels associated with adverse effects on human health and the sustainability and health of biota?
- Are observed changes in exposure attributable to mitigation measures or changes influenced by the Convention?

6) Quantification of key environmental processes to improve our understanding of cause-effect relationships

- What does the level, spatial pattern, or temporal trends of mercury suggest about the relative importance of environmental processes and parameters driving transport and fate?
- How consistent are the observed levels, spatial patterns and temporal trends with the modelled estimates and what lessons can be learned from them to improve the existing models?

Different types of observations and sources of data may be most appropriate for addressing different questions during the Effectiveness Evaluation. The more basic scientific questions on levels and trends of mercury in humans and the environment are often more easily answered with a high level of confidence than the more complex questions related to source attribution and exposure assessment. However, the more complex questions are more accurately associated with specific measures under the Convention. How confidently and accurately a particular question can be answered will, in turn, depend on the quality and representativity of the available data, and the robustness of the scientific analysis. Together, the operational questions, and the confidence and accuracy by which they can be answered form a successive continuum of scientific weight-of-evidence that will help us understand the effectiveness of the Convention.

Air, biota and humans have been identified as key matrices for monitoring trends in the movement of mercury from its sources to the environment and into human populations. Designing a monitoring strategy to link observed changes in these matrices will be important to evaluate the effectiveness of the Convention. National and multi-country monitoring programmes, including those identified in the Supplementary Material, as well as programmes and projects overseen by international organizations, may be prioritized as the preferred sources of comparable monitoring data. In the absence of those, additional sources of comparable data may also provide valuable information to support the Effectiveness Evaluation. Quality control measures, including those listed in chapters 3-5 and Supplementary Material, will be needed to assess the usefulness and validity of different data sets, and maximize scientific weight of evidence.
A tiered approach to monitor trends in the movement of mercury and mercury compounds in the different matrices is put forward in this guidance with a view to supporting Parties and organizations who wish to develop new monitoring programs or improve existing ones. The approach is comprised of three tiers:

**Tier 1** is intended to provide guidance on mercury monitoring under a limited set of parameters for circumstances where available resources are not sufficient to implement the actions in Tier 2. Following guidance by the COP,¹ the methods in Tier 1 are cost effective, practical, feasible, and sustainable. The Tier 1 methods are intended to provide information that are useful in identifying and characterizing gaps and needs of national, regional, or local interest and to provide information that is useful to the collective effort for the Effectiveness Evaluation. While the implementation of Tier 1 actions may not fully address the questions in Table 2.1, it will contribute essential information and create a foundation for Tier 2 monitoring.

**Tier 2** is intended to build upon Tier 1 methods to provide information that will address the questions identified in Table 2.1, and to create a basis for assessing source attribution at the local, national, and global scales. The methods and approaches in this tier may be more expensive or complex than those under Tier 1. The more comparable data from Tier 2 becomes available, the more robust the Effectiveness Evaluation will be.

**Tier 3** identifies research methods and approaches that may play a vital role in supporting the Tier 1 and Tier 2 programs and the Effectiveness Evaluation, primarily by improving our understanding of key processes that link sources to environmental concentrations and exposures. Because Tier 3 focuses on processes, the results would likely yield insights that are broadly applicable and that should be taken into consideration in the Effectiveness Evaluation when available.

The tiered approach is further elaborated in the annex with an overview of what data are collected under each tier for the three monitoring matrices.

**Atmospheric mercury monitoring**

Mercury levels in the atmosphere are linked to mercury emissions from natural and anthropogenic sources. Key anthropogenic sources of atmospheric mercury include point sources listed in Annex D of the Convention, the intentional use of mercury in artisanal and small-scale gold mining (ASGM), and in certain industrial products and processes. In the context of the effective evaluation, it will be relevant to estimate how large the contribution of sources in Annex D is as compared to total anthropogenic emissions. Many of the Convention measures to control mercury supply, use, emissions, storage, and disposal are expected to reduce levels of mercury in the atmosphere.

Chapter 3 identifies different methods Parties and organizations can use to monitor atmospheric mercury and generate comparable data to support the Effectiveness Evaluation of the Convention.

Monitoring of atmospheric mercury has been ongoing for decades but not all regions are equally covered with the biggest data gaps in the southern hemisphere. A tiered approach is proposed and gives Parties and organizations an opportunity to start, expand or improve their monitoring programmes in such a manner that comparable data can be generated to support the Effectiveness Evaluation.

¹ Decision MC-2/10 pursuant to the terms of reference to Ad-hoc Technical Expert Group on Effectiveness Evaluation.
Moreover, the tiered approach also breaks down the monitoring requirements in such a manner that new atmospheric mercury monitoring initiatives have an opportunity of joining one of the several existing monitoring programmes and networks, thus drawing from the experience and information at hand that these established networks can provide.

Automated atmospheric mercury measurement is the preferred method within existing monitoring networks. While the instruments used in automated measurements are capable of detecting very low concentrations of mercury, these instruments are expensive and alternative options are available that can also deliver comparable data. Passive and manual sampling of atmospheric mercury are two such options, even if at a lower temporal resolution as compared to automated systems.

Depending on the specific needs of the monitoring initiative, this guidance puts forward different methods at Tier 1, as the minimum step to start generating comparable atmospheric mercury data. Wet deposition of Hg from the atmosphere is one of the methods included at Tier 1 level. The method is reasonably well understood and sufficient results have been achieved in networks, as well as on a global scale, through various studies and intercomparison exercises. Therefore, scientifically sound, and cost-effective methods and techniques to determine mercury concentrations in air are available and can generate comparable data.

Another important factor to take into account when performing mercury air monitoring is the location(s) where monitoring will take place. Monitoring at a variety of sites will provide a more comprehensive picture of the levels of Hg in the atmosphere. It is therefore important for each monitoring initiative to identify sites that can provide insights into changes in atmospheric mercury levels over time, including relevant and sensitive ecosystems. Carefully selected sites can also help develop more robust atmospheric models and fill data gaps.

Beyond sampling and analysis, for any monitoring programme to be successful, a strong quality assurance and quality control (QA/QC) program is needed. A wealth of experience on key elements and processes related to QA/QC is available from existing atmospheric mercury monitoring programmes and networks as seen in this chapter and Supplementary Material.

The elements put forward in chapter 3 will help answer various operational questions for the Effectiveness Evaluation with regard to atmospheric mercury monitoring. Furthermore, chapter 3 provides Parties and relevant organizations with the means of starting, improving or expanding on their initiatives for monitoring atmospheric mercury to enable them to deliver comparable data for the Effectiveness Evaluation.

**Biota mercury monitoring**

Mercury released into the environment may be converted to other forms, such as methylmercury, and accumulated in fish and wildlife, and it can negatively impact fish, wildlife and human health through the consumption of contaminated prey and food. Mercury may also cause significant negative impacts to the environment, for example by adversely impacting ecosystem services and leading to the loss of rare species and potentially biodiversity. Article 22 of the Convention requires the Conference of the Parties to establish arrangements to provide monitoring data on the trends in levels of mercury observed in biotic media.

Data from monitoring mercury concentrations in biota can help address the operational questions for the Effectiveness Evaluation of the Convention. Historic data available from various biota monitoring
programmes, databases and other resources can be used to improve our understanding of the
exposures to mercury in biota before the Minamata Convention’s entry into force and to help establish
a baseline for the Effectiveness Evaluation. Moving forward, existing government-led national mercury
monitoring programs, regional initiatives, and/or academic-led studies can provide comparable biota
monitoring data for use in the Effectiveness Evaluation. New monitoring efforts may further contribute
by providing comparable data on key bioindicators filling data gaps, and building capacity.

Biota monitoring data can be collected with focus on continental and oceanic components designed as
part of a Tiered approach for Parties and organizations who elect to develop new monitoring programs
or improve existing ones. Briefly, Tier 1 is suitable for initiatives seeking to create a biota-based
monitoring programme, or expand a minimal programme, but that may not have sufficient resources to
implement the actions in Tier 2. Tier 1 focuses on identifying temporal trends and collecting total
mercury measurements from trophic level 3 or 4 biota that best represent the targeted habitats. Tier 1
activities should ideally be repeated for the same species using the same size classes in the same habitat
every 2-5 years. Tier 2 aims to generate information that will contribute more meaningfully to the
Effectiveness Evaluation, and thus calls for more in-depth analysis of the Tier 1 monitoring efforts, or
incorporation of mercury biomonitoring into other, in-depth mercury monitoring efforts. Tier 3 aims to
increase understanding of key processes that link mercury sources to key bioindicators, and thus
resource-intensive research methods and approaches are required.

Chapter 4 describes the key elements that are essential to all monitoring efforts for biota, including: a)
defining the target bioindicators and sample size, which usually focus on high trophic level biota that are
vulnerable to relatively high methylmercury exposure; b) selecting and measuring the appropriate
biomarkers (i.e. tissue types) to best interpret exposure to different sources and forms of mercury, with
total mercury measurements in muscle tissue of fish and marine mammals, blood, and feathers or eggs
of birds being most commonly used and accepted; c) identifying the monitoring locations that best
reflect the objective for biomonitoring (e.g., temporal, spatial, or ecological health questions) through
the use of an ecosystem sensitivity modelling tool; and d) managing and analysing data as per the
guiding operational questions for the Effectiveness Evaluation. All these aspects can use well-established
standard operating procedures available in the Supplementary Material.

Human biomonitoring

Human health may be negatively impacted by mercury exposure. Human populations may be exposed
to elemental and inorganic mercury in occupational settings (e.g. in ASGM and dentistry), from contact
with certain products (e.g. dental amalgams, some skin-lightening creams, broken fluorescent bulbs and
other waste products), and from environmental contamination, as well as to organic mercury largely
from dietary sources, including but not limited to shellfish, fish, and marine mammals contaminated
with methylmercury. Measuring mercury levels in the blood, hair and/or urine of individuals from target
populations provides direct information on human exposures to mercury, from which risks to human
health can be assessed.

Article 22 of the Convention requires the Conference of the Parties to establish arrangements to provide
monitoring data on the trend in mercury levels in vulnerable populations, and human biomonitoring
data can help address the operational questions that will support the Effectiveness Evaluation.
Chapter 5 provides essential guidance, and links to key resources, for Parties and relevant organizations
human biomonitoring data for the Effectiveness Evaluation.

In terms of using existing biomonitoring data, several databases and resources exist, and these can be used to help understand human exposures to mercury before the Minamata Convention’s entry into force and help establish the baseline for the Effectiveness Evaluation. In terms of data to be realized in the future, there are two sources to consider. First, biomonitoring data generated by existing government-led national biomonitoring programs, regional initiatives, and/or academic-led studies. Second, Parties and relevant organizations can further support the Effectiveness Evaluation by implementing new biomonitoring studies in a harmonized way so that they are purposefully designed to fill data gaps, and build capacity.

Human biomonitoring data can be designed as part of a tiered approach to inform new monitoring programmes or improve existing ones. Briefly, Tier 1 is appropriate for initiatives seeking to create a human biomonitoring programme, or expand a minimal programme, but that may not have sufficient resources to implement the actions in Tier 2. The goal of Tier 1 is to focus on a vulnerable sub-population and take total mercury measurements in blood, urine, or hair. This activity should ideally be repeated in the same population every 2-5 years. A good starting point for Tier 1 is the recent guidance from the WHO to characterize prenatal mercury exposure (WHO 2018a). Tier 2 aims to realize information that will help address all operational questions for the Effectiveness Evaluation, and thus calls for more in-depth analysis of the Tier 1 sub-population group, or incorporation of mercury biomonitoring into other, in-depth health surveys or cohort studies. Tier 3 aims to increase understanding of key processes that link mercury sources to human exposures, and thus resource-intensive research methods and approaches are required.

Key elements to all human biomonitoring studies that need to be considered, include: a) defining the target and sample population (which usually focus on groups vulnerable to mercury i.e. early lifestages or those with relatively high exposures); b) selecting and measuring the appropriate biomarkers to help tease apart exposure to different sources and forms of mercury (with total mercury measurements in hair, urine, blood and cord blood being most commonly used and accepted); c) administering surveys to gather supportive information (e.g., on socio-demographics, occupational practices, dietary habits) to deepen understanding; and d) managing and analyzing data as per the guiding policy question. All these aspects must be performed in a responsible and ethical manner.

**Cross-media data management, modelling and analysis**

Because the connections between monitoring media are not necessarily direct and instantaneous but do depend largely on known or suspected physical processes, mechanistic models explicitly representing these processes are a valuable tool for interpretation of monitoring results and can thereby contribute to the Effectiveness Evaluation. However, as the complexity of the data increases, identifying all the relevant processes and quantifying them correctly becomes more challenging. In such cases, mechanistic models can be supplemented with different kinds of statistical models.

From primary release to human exposure, mercury can undergo many physical and (bio-)chemical changes which interact with each other over a large range of timescales and can be influenced by human behaviour. Attribution of observed trends to specific drivers such as direct anthropogenic mercury releases, legacy mercury, natural process-driven releases, and non-mercury environmental or behavioural drivers requires the use of models which resolve the intervening processes supplemented...
or calibrated by empirical statistical approaches. This makes cross-media analysis involving both mechanistic and statistical modelling in all relevant media an important part of the weight of evidence useful to evaluate effectiveness of the Convention.

By analysing monitoring data, temporal and spatial trends in the levels of mercury in specific environmental media or human matrices can be derived. These trends provide a first-level indication of whether the Convention may be contributing to protecting human health and the environment from the adverse effects of mercury. Analyses of the monitoring data collected in each matrix separately will be informative, and these monitoring data can also be used in an integrated manner, where combining multiple complementary analysis approaches to answer the same question will improve robustness and increase the scientific weight of evidence. This is particularly important when models are used for policy evaluation applications and uncertainties need to be quantified and minimized.

In many cases, attribution of observed trends to specific drivers can be performed through the use of models which resolve the intervening processes, supplemented by empirical statistical approaches. Cross-media analysis involving both mechanistic and statistical modelling in all relevant media is important in order to fully evaluate effectiveness of the Convention. This evaluation requires separating the impacts attributable to the Convention from changes that occur due to other factors, and while monitoring data shows the impact of all of these factors, modelling can help attribute the changes to the different drivers. As more monitoring data and analysis tools become available, more detailed analysis can be performed.

To estimate levels of mercury in locations with or without anthropogenic mercury sources, simple analyses can be conducted on monitoring data at sites chosen for this purpose. Temporal trends can be identified at these and other locations once a long enough time record has been collected. This trend analysis should account for variability and uncertainty to obtain trend magnitudes, confidence intervals for the trends, and measures of the trends’ statistical significance.

To characterize spatial patterns, several atmospheric chemical transport models can be used, supplemented with statistical models where beneficial to quantify representativeness of observed levels and trends in air, and to extrapolate ambient air concentrations and wet deposition to areas with sparse monitoring data. Spatially resolved models in air and other media can be used to interpolate levels and trends of mercury while accounting for the drivers of spatial and temporal differences.

Both “bottom-up”, or process-based analyses that estimate effects of drivers on observable quantities, and “top-down”, or observation-based analyses for identification/estimation of drivers, types of analyses are valuable to modelling. Bottom-up analyses can be performed with atmospheric models for source attribution, and top-down analyses for air and biota attribution where sufficient ancillary data is available. Top-down analysis of changes in exposure pathways can also be performed to attribute changes in human biomarkers to measures influenced by the Convention. At intensive monitoring sites, combined top-down and bottom-up attribution analyses can be performed for air, biota and human biomarkers. To quantify legacy impacts, coupled-media approaches should be used, where possible.

Exposure can be estimated based on specific sources and exposure attribution information can be used to estimate marginal health impacts/costs of individual drivers. Trends in risk associated with trends in exposure and/or biomarker benchmark values can be estimated where the appropriate information is available.
The quantification of key environmental processes can improve our understanding of cause-effect relationships. Top-down analysis can be used to identify key environmental drivers, and large-scale measurement/model intercomparisons can be performed to identify key processes. Improved understanding can lead to a beneficial iterative approach: using the available information to improve the application of models can decrease the uncertainty for further and future analyses and evaluations.

Further to the present main document, this guidance also has a Supplementary Material organised in two parts – Part A contains an overview of existing monitoring programmes, organised per matrix (air, biota and human biomonitoring), and an overview of existing gaps, as well as a non-exhaustive list of standard operating procedures; and Part B contains an overview of quality assurance and quality control procedures in laboratory analysis and data management, and includes a draft template for the submission of monitoring data.
Chapter 1. Introduction and Objectives

1.1. Introduction

The objective of the Minamata Convention on Mercury (herein referred to as the Convention) is to protect the human health and the environment from anthropogenic emissions and releases of mercury and mercury compounds (Article 1). The Convention contains, in support of this objective, provisions that relate to the entire life cycle of mercury, including controls on the supply and trade of mercury, products and processes where mercury is used, emissions and releases of mercury, and management of waste and contaminated sites (Articles 3-12). The Convention also includes provisions that support the Parties to fulfil their obligations (Articles 13 and 14), health aspects (Article 16), and measures to enhance knowledge and information (Articles 17-19). Article 22 of the Convention requires the Conference of the Parties (COP) to periodically evaluate the effectiveness of the Convention, and to perform this evaluation on the basis of available scientific, environmental, technical, financial and economic information. Comparable monitoring data on the presence and movement of mercury and mercury compounds in the environment, as well as trends in levels of mercury and mercury compounds observed in biotic media and vulnerable human populations, are of particular interest to COP in the context of the Effectiveness Evaluation.

1.2. Objectives

This document, as requested by the COP in its decision MC-3/10 in November 2019, provides scientific and technical guidance to support the COP to obtain comparable monitoring data for the Effectiveness Evaluation. The primary objectives of this document are to:

- Explain the role of monitoring in the Effectiveness Evaluation and set realistic expectations about what can be learned over time.
- Provide guidance to Parties and organizations, which are currently conducting monitoring programmes, on what data and accompanying information would inform the Effectiveness Evaluation.
- Provide guidance to Parties and organizations who wish to develop new monitoring programs, or improve existing ones, with a view to contributing to the Effectiveness Evaluation.

This document describes the scientific and technical processes and guiding principles for compiling and/or generating comparable monitoring data, as well as methods to use such monitoring data for understanding the presence, movements and trends of mercury in the environment and humans, in the context of evaluating effectiveness of the Convention.

Chapter 2 builds on the four overarching policy questions proposed for the Effectiveness Evaluation and establishes five categories of monitoring activities that can produce comparable data to address these questions. It explains the rationale for selecting air, biota and human as core matrices for monitoring activities, and presents general guidance that is relevant to all matrices to support efforts towards obtaining comparable monitoring data.

Following chapters address monitoring of mercury in specific matrices: air (chapter 3), biota (chapter 4) and humans (chapter 5). These chapters describe the significance of monitoring the matrices, and provide guidance on the selection of monitoring sites, sampling and measurement methods, quality control and assurance, and data collection, management, analysis and evaluation.
Chapter 6 discusses how these matrix-specific data can be compiled, analysed and synthesized, how those data can be used in mechanistic and statistical models, and how observed changes in mercury levels in environmental media and humans observed can be interpreted.

The Annex contains a proposed tiered approach for programmes to monitor mercury and mercury compounds to support the Effectiveness Evaluation.

The Supplementary material to the guidance presents an overview of existing monitoring activities undertaken by Parties and other stakeholders, as well as a review of gaps in the monitoring of key matrices. The Supplementary Material will be a “living document” that may be updated to support the COP in identifying available monitoring information for the Effectiveness Evaluation, as well as to support Parties and relevant organizations to consider whether their monitoring activities could contribute to filling the gaps. Other supplemental information will be developed to support the use of this document, including the comparison of existing standard operating procedures, international QA/QC programmes, and available reference materials.
Chapter 2. Comparable Monitoring Data and the Effectiveness Evaluation

2.1. Introduction

Paragraph 2 of Article 22 on Effectiveness Evaluation of the Minamata Convention requires the Conference of the Parties (COP) to make “arrangements for providing itself with comparable monitoring data on the presence and movement of mercury and mercury compounds in the environment as well as trends in levels of mercury and mercury compounds observed in biotic media and vulnerable populations”. It has been proposed that the Effectiveness Evaluation of the Convention should address the following four overarching policy questions:

1) Have the Parties taken actions to implement the Minamata Convention?
2) Have the actions taken resulted in changes in mercury supply, use, emissions and releases into the environment?
3) Have those changes resulted in changes in levels of mercury in the environment, biotic media and vulnerable populations that can be attributed to the Minamata Convention?
4) To what extent are existing measures under the Minamata Convention meeting the objective of protecting human health and the environment from mercury?

Monitoring levels of mercury in air, biota, and humans can contribute to addressing the third and fourth policy questions above. Detecting changes in mercury levels, estimating the human or ecosystem health impacts of those changes, and attributing them to actions influenced by the Minamata Convention require the use of mechanistic and/or statistical models. Therefore, observations are needed not only to detect and quantify changes, but also to improve and evaluate models of mercury transport, fate, exposure, and impacts.

The overarching policy questions can be operationalized through monitoring activities grouped into five categories:

1) Estimation of contemporary mercury concentrations for areas without (i.e. background sites) or with (i.e. affected sites) local anthropogenic sources;
2) Identification of temporal trends;
3) Characterization of spatial patterns;
4) Estimation of source attribution;
5) Estimation of exposure and adverse impacts;
6) Quantification of key environmental processes to improve our understanding of cause-effect relationships.

---

2 Document UNEP/MC/COP.3/14 and further information therein.
2.2. Weight of evidence and operational questions

From each of the monitoring categories above, operational questions can be drawn, as outlined in Table 2.1, to guide the collection and analysis of the relevant monitoring data and inform the Effectiveness Evaluation in complementary ways. Different types of observations and sources of data may be most appropriate for addressing different questions during the Effectiveness Evaluation.

National and multi-country monitoring programmes, including those identified in the Supplementary Material, as well as programmes and projects overseen by international organizations, such as WHO and the GEF, may be prioritized as the preferred sources of comparable monitoring data. In the absence of those, additional sources of comparable data may also provide valuable information to support the Effectiveness Evaluation. Quality control measures, including those listed in chapters 3-5 and Supplementary Material, will be needed to assess the usefulness and validity of different data sets and maximize scientific weight of evidence.

The more basic scientific questions on levels and trends of mercury in humans and the environment are often more easily answered with a high level of confidence than the more complex questions related to source attribution and exposure assessment. However, the more complex questions are more accurately associated with specific measures under the Convention. How confidently and accurately a particular question can be answered will, in turn, depend on the quality and representativity of the available data, and the robustness of the scientific analysis. Together, the operational questions in Table 2.1, and the confidence and accuracy by which they can be answered form a successive continuum of scientific weight-of-evidence that will help us understand the effectiveness of the Convention.

Starting with the most basic scientific question, i.e. comparing the levels of mercury in background and impacted locations to established benchmark values, it is relatively easily done with a high level of confidence. When discussed together with other available information from the EE process, these mercury levels alone give valuable information for the Effectiveness Evaluation, even if the scientific link to specific measures influenced by the Convention is weak. Identifying possible temporal trends and spatial patterns adds further to the weight-of-evidence, because impacted sites can be expected to react faster than background sites to measures influenced by the Convention, even if no formal cause-effect relationships is established. The levels of mercury and the identification of temporal trends can be determined with the sampling strategies and analytical methods (see Tier 1 below).

Estimating exposure and adverse impacts from mercury in habitats, ecosystems or populations will further help to understand the effectiveness of the Convention. This information is independently valuable for the Effectiveness Evaluation, but when combined with the methods used to estimate source attribution, the full pathway from source to impact can be described. By conducting a formal analysis that can estimate what sources are causing changes in levels of mercury in humans or the environment, the weight of evidence can be further increased, compared with just describing temporal trends or spatial patterns. Statistical methods can be used to infer relationships between observed mercury levels and potential drivers.

Mechanistic models that represent physical processes can be used to examine the consistency of these inferred relationships with what is known about other processes and to estimate mercury levels in the environment. These two methodological approaches can be used separately, but the strongest weight of evidence is obtained when they are used together in a complimentary manner. How accurately levels of mercury in humans and the environment can be attributed to changes in specific sources will depend on
the available data, and the robustness and confidence of the used model. For example, it is easier to estimate the relative contribution of primary anthropogenic sources from natural and legacy mercury, compared with estimating accurately the absolute contribution of specific sources influenced by the Convention.

The continuum of scientific weight of evidence that the operational questions in Table 2.1 provide is reflected in the three tiers presented in this guidance. Together, answers to the operational questions form a successive continuum of scientific weight of evidence that can form an evidentiary basis for addressing policy questions 3 and 4 as part of the Effectiveness Evaluation.

**Table 2.1 Monitoring activities and associated operational questions**

1) Estimation of background and impacted levels of mercury
   - What are the levels and form of mercury in sites that are expected to be remote from anthropogenic sources?
   - What are the levels and form of mercury in sites that are expected to be affected by local anthropogenic point sources?

2) Identification of temporal trends
   - Do the levels and form of mercury in the observed matrix (air, biota, human) at a given location change over time – for example, in short- (< 5 years), medium- (5 to 20 years), and long-term periods (> 20 years)? Is there a long-term trend or trajectory (a signal) that can be separated from the temporal variability (noise)?
   - How do observed temporal variations and trends differ spatially?
   - How do observed temporal variations and trends in mercury compare to or co-vary with variations and trends of:
     - Mercury in different forms (chemical species) or other within matrices?
     - Mercury emissions and releases?
     - Related pollutants/emissions or environmental variables?

3) Characterization of spatial patterns
   - What are the levels and form of mercury in the observed matrix (air, biota, human) at a given location and time?
   - Taken together, what does the available data suggest about:
     - Spatial variability in environmental mercury concentrations?
     - Variability in environmental mercury concentrations within populations, their habitats, and ecosystems?
   - Do the observed spatial variations and patterns or gradients differ among:
     - Forms (chemical species) of mercury?
     - Environmental matrices?
   - How do the observed spatial variations and patterns or gradients compare to those of:
     - Mercury emissions and releases?
     - Related pollutants/emissions or environmental variables?

4) Estimation of source attribution
• Using models and statistical analyses consistent with observational data, how can the observed levels, spatial patterns, temporal trends and adverse impacts on species, ecosystem services, biodiversity and human populations be attributed to changes:
  ➢ In sources or sinks of anthropogenic, natural and legacy mercury?
  ➢ In anthropogenic sources (local, regional, global) of mercury?
  ➢ Influenced by the Convention?
• Not influenced by the Convention?

5) Estimation of exposure and adverse impacts
• How do mercury concentrations impact populations, species, ecosystem services, biodiversity and human populations in regions that are remote from sources as well as those that are locally impacted by anthropogenic sources?
• How do the observed levels of mercury in air, biota, and humans compare to established national and international benchmark levels associated with adverse effects on human health and the sustainability and health of biota?
• Are observed changes in exposure attributable to mitigation measures or changes influenced by the Convention?

6) Quantification of key environmental processes to improve our understanding of cause-effect relationships
• What does the level, spatial pattern, or temporal trends of mercury suggest about the relative importance of environmental processes and parameters driving transport and fate?
• How consistent are the observed levels, spatial patterns and temporal trends with the modelled estimates and what lessons can be learned from them to improve the existing models?

2.3. Monitoring matrices

Mercury is a chemical of global concern owing to its long-range atmospheric transport, persistence in the environment, and ability to biomagnify and bioaccumulate in ecosystems leading to adverse effects on human health and the environment. For the purpose of evaluating the effectiveness of the Convention, in the light of its objective (i.e., “to protect the human health and the environment from anthropogenic emissions and releases of mercury and mercury compounds”), it is important to monitor temporal and spatial changes in the movement of mercury from its sources to the environment and into human populations. As such, air, biota and humans have been identified as the key matrices for tracking mercury (Figure 2.1) and designing a monitoring strategy to link observed changes in these matrices will be important to evaluate the effectiveness of the Convention.

Air: Mercury levels in the atmosphere are linked to mercury emissions from natural and anthropogenic sources. Key anthropogenic sources of atmospheric mercury include point sources listed in Annex D of the Convention, the intentional use of mercury in artisanal and small-scale gold mining (ASGM), and in certain industrial products and processes. In the context of the effective evaluation, it will be relevant to estimate how large the contribution of sources in Annex D is as compared to total anthropogenic emissions. Many of the Convention measures to control mercury supply, use, emissions, storage, and disposal are expected to reduce levels of mercury in the atmosphere.
Biota: Mercury released into the environment may be converted to other forms, such as methylmercury, and accumulated in fish and wildlife, and it can negatively impact fish, wildlife and human health through the consumption of contaminated prey and food. Mercury may also cause significant negative impacts to the environment, for example by adversely impacting ecosystem services and leading to the loss of rare species and potentially biodiversity. Article 22 of the Convention requires the COP to establish arrangements to provide monitoring data on the trends in levels of mercury observed in biotic media.

Human biomonitoring: Human health may be negatively impacted by mercury exposure. Human populations may be exposed to (a) elemental and inorganic mercury in occupational settings (e.g. in ASGM and dentistry), from contact with certain products (e.g. dental amalgams, some skin-lightening creams, broken fluorescent bulbs and other waste products), and from environmental contamination; and (b) organic mercury largely from dietary sources (including but not limited to shellfish, fish, and marine mammals contaminated with methylmercury). Human biomonitoring (i.e., measuring mercury levels in the blood, hair and/or urine of individuals from a target population, depending on the form of mercury exposure) provides direct information on human exposures to mercury, from which risks to human health can be assessed. Article 22 of the Convention requires the COP to establish arrangements to provide monitoring data on the trend in mercury levels in vulnerable populations.

Other matrices: Available monitoring data for environmental matrices such as freshwater, sediments, vegetation, snowpacks, soils, and oceans may also be useful, in certain contexts, to support the Effectiveness Evaluation. Levels of mercury in freshwater may be helpful to assess environmental contamination in a given area. However, due to the complexity of tracking mercury contamination in biota resulting from water contamination, direct measurements of mercury concentrations accumulated in fish, marine mammals, sea turtles and birds offer a more practical indicator for assessing environmental contamination. Soil and sediment monitoring can provide data for an assessment of the local environment, especially in heavily polluted areas, and information from such monitoring may be used to inform the effectiveness of specific measures in certain areas, such as those addressing ASGM. As the commercial seafood market is the largest source of human exposure to MeHg on a global basis, levels of mercury in the oceans’ surface may also contribute to the assessment of global environmental transport of mercury.

Based on these considerations, this document provides guidance on air monitoring, biota monitoring and human biomonitoring. Examples of ancillary measurements from water and sediment samples associated with biota monitoring are included in chapter 4. Review of environmental monitoring methods that may be used to assess ASGM sites is also being developed to support the implementation of Article 7 of the Convention. An overview of existing monitoring programs is provided in the Supplementary Material to this guidance.

---

3 The Secretariat is developing a separate document which will be made available for consideration by the fourth meeting of the COP.
2.4. Tiered approach for developing and improving monitoring programmes

To support Parties and organizations who may wish to develop new monitoring programs, or improve existing ones, with a view to contributing to the Effectiveness Evaluation, this document identifies a tiered approach for monitoring each of the three media (air, biota, humans):\textsuperscript{4}

- **Tier 1** is intended to provide guidance on mercury monitoring under a limited set of parameters for circumstances where available resources are not sufficient to implement the actions in Tier 2. Following guidance by the COP,\textsuperscript{5} the methods in Tier 1 are cost-effective, practical, feasible, and sustainable. The Tier 1 methods are intended to provide information that are useful in identifying and characterizing gaps and needs of national, regional, or local interest and to provide information that is useful to the collective effort for the Effectiveness Evaluation. While the implementation of Tier 1 actions may not fully address the questions in Table 2.1, it will contribute essential information and create a foundation for Tier 2 monitoring.

- **Tier 2** is intended to build upon Tier 1 methods to provide information that will address the questions identified in Table 2.1, and to create a basis for assessing source attribution at the local, national, and global scales (Figure 2.2). The methods and approaches in this tier may be

---

\textsuperscript{4} It is noted that the Convention does not impose any obligation upon Parties to conduct monitoring. As such, the tiered approach and any other activities or recommendations contained in this guidance are voluntary and presented with the sole purpose of supporting Parties who may wish to develop new monitoring programs, or improve existing ones, with a view to contributing to the Effectiveness Evaluation.

\textsuperscript{5} Decision MC-2/10 pursuant to the terms of reference to Ad-hoc Technical Expert Group on Effectiveness Evaluation.
more expensive or complex than those under Tier 1. The more comparable data from Tier 2 becomes available, the more robust the Effectiveness Evaluation will be.

- **Tier 3** identifies research methods and approaches that may play a vital role in supporting the Tier 1 and Tier 2 programs and the Effectiveness Evaluation, primarily by improving our understanding of key processes that link sources to environmental concentrations and exposures. Because Tier 3 focuses on processes, the results would likely yield insights that are broadly applicable and that should be taken into consideration in the Effectiveness Evaluation when available.

**Figure 2.2.** Illustration of attribution across matrices for hypothetical contributions of selected drivers at a hypothetical location. The coloured bars represent the fractional contributions of different drivers to observed mercury trends/variability in each medium. The drivers of variability/change in a given medium can in turn be attributable to drivers in other media (C. Thackray, unpublished).

### 2.5. Quality of monitoring data

The Effectiveness Evaluation of the Convention will require monitoring data that is comparable and credible. The Quality Assurance / Quality Control (QA/QC) protocols employed by existing monitoring programs will provide a basis to inform the development of comparable data for use in the Effectiveness Evaluation. Data generated from different monitoring programs may be supplemented, as appropriate, with comparable data from academia and research. This may be accomplished through a well-documented and transparent set of "data flags" that will enable the use of data from different sources with different levels of QA/QC.

Understanding the presence and movement of mercury in the environment and in humans will require implementation of different monitoring programmes that yield good quality data via QA/QC protocols. Such programmes need to be well documented and capable to integrate data from across different monitoring programmes for the purposes of the Effectiveness Evaluation.

Examples of criteria to assess the quality of mercury monitoring data include:
• Selection bias – describe the location/population/setting (Is sampling performed in a consistent manner, representative of the target location/population/setting, and bias-free?);

• Exposure detection – describe how mercury was measured in a given sample (Is the measurement value accurate and precise, and does it follow a standardized and scientifically credible method? Are additional metadata provided to do exposure/attribution assessment?);

• Statistical parameters – describe the sample size and its adequacy, and whether basic and essential data is present, complete and well summarized (Is the data useful to address the questions in Table 2.1?).

In order to determine adequate sampling site location, sample sizes, site densities, sampling frequencies, or observational period needed to detect expected changes in the context of the Effectiveness Evaluation, an analysis based on the spatial and temporal variability of existing data and model projections of potential changes may be necessary. Chapters 3, 4 and 5 describe specific QA/QC considerations for each medium or matrix.6

2.6. Data management

To the extent possible and in accordance with requirements of individual data providers, data used in the Effectiveness Evaluation should follow the FAIR principles (findable, accessible, interoperable and reusable)7 for data management and stewardship, including the following elements:

Findable:

• A searchable and interoperable database acting as a repository of available data;

• Unique identification systems (e.g. “Digital Object Identifiers” or “DOIs”) and controlled vocabulary to facilitate searching and retrieval of information;

• Detailed metadata associated with each data record to facilitate the submission, searching, location and retrieval of information;

Accessible:

• Free and open access to the data to Governments, Indigenous Peoples, and relevant stakeholders, taking into account the relevant ethical considerations;

Interoperable:

• An interoperability mechanism to facilitate the exchange of information across different programmes and databases;

Reusable:

• Data usage license/agreement identifying the terms and conditions for further use of the data;

• Metadata including enough information describing how the data were collected/produced to enable an assessment of the quality and comparability of the data, reproducibility and further analyses.

Further to the FAIR principles to facilitate increased data sharing, principles to support an ethical use of data should also be followed (see chapter 5). Ethical considerations associated with Indigenous Peoples,

---

6 The terms “medium/media” and “matrix/matrices” are being used interchangeably throughout this document.

7 Wilkinson et al. (2016).
including with regards to self-determination of research, research ethics, data considerations, utilization of Indigenous Knowledge, and communication of results, should be guided by principles such as the “CARE Principles for Indigenous Data Governance”.  

Major international and national monitoring programmes and networks have their own data management systems, including data repositories, portals, or catalogues. Some of these may be used as primary repositories, along with other sources of data, including Indigenous Knowledge in mercury monitoring efforts, to gather and exchange information so that monitoring data, relevant metadata, ancillary data and QA/QC information from different programs can be used for the Effectiveness Evaluation. There are also global initiatives to develop monitoring data management platforms that enable access to data from multiple monitoring programmes, networks and existing primary data repositories. An overview of the existing data management systems and networks will be provided in a Supplementary Material to this guidance.

---

Chapter 3. Atmospheric Mercury Monitoring

3.1. Introduction

Monitoring of atmospheric mercury has been identified as one of the primary and most appropriate types of monitoring to assist the Convention in determining its effectiveness (AMAP 2011; Evers et al., 2016; Gustin et al. 2016; Sprovieri et al. 2016; UNEP 2018). Monitoring of atmospheric mercury generally has the following objectives: 1) to gather information on spatial and temporal trends of changes in mercury concentrations in the atmosphere; 2) assessing atmospheric mercury inputs to aquatic and terrestrial ecosystem and 3) to provide data for the development and improvement of transport and chemistry models (Obrist et al., 2018; Saiz-Lopez et al. 2020; Skov et al. 2020). Monitoring may also provide insight on how climate change affects, and is affected, by atmospheric mercury.

This chapter aims to provide guidance to Parties and relevant organizations on the different methods to monitor mercury in air, and what vital procedures need to be in place to continue, expand or start with monitoring mercury in air to generate comparable monitoring data to support the Effectiveness Evaluation. Given that the development and operationalization stages differ across countries and programmes, a tiered approach offers a way forward to support Parties and relevant organizations in starting or expanding on their monitoring programme (see section 3.3 and annex to this document).

3.2. Significance of air as a matrix for mercury monitoring

Mercury is a naturally occurring element and is emitted to the atmosphere from a variety of natural geogenic sources, legacy emissions and anthropogenic sources (Driscoll et al., 2013). Atmospheric deposition represents the major pathway of mercury input to terrestrial and aquatic ecosystems outside areas with substantial sources that directly release mercury to land or water. Moreover, mercury that has been deposited on land and ocean surfaces can be re-emitted to the atmosphere, through the so-called “legacy emissions”. Monitoring efforts and modelling studies have shown that the atmosphere responds relatively quickly, i.e. days to months, (though not proportionally due to legacy emissions and natural sources) to decreasing mercury emissions (Layman et al., 2020). Mercury is present in the atmosphere as gaseous elemental mercury (Hg⁰ or GEM) or as reactive mercury species (HgII) in the gas or particle phase. The relaxation time for GEM in the atmosphere is relatively long — approximately 12 months compared to the other forms (gaseous oxidized mercury and particle-bound mercury) with a short residence time in the atmosphere ranging from a few hours to weeks (Horowitz et al., 2017; Skov et al. 2020). This, together with the influence of re-emissions of previously deposited mercury (and the presence of geogenic sources), means that the change in atmospheric concentrations or deposition will be less than proportional to the change in primary anthropogenic emissions and there will be a delay between abatement and observation of a response in atmospheric mercury levels (Sprovieri et al., 2016). The lag time is expected to be much shorter than the time response for mercury other reservoirs (soils, surface waters, biota and ocean) where residence times are much longer (approximately decades) and mercury levels are complicated by several other factors (Lyman et al., 2020). Therefore, at large regional scales, the atmosphere is expected to be one media where changes in environmental levels due to changes in emissions influenced by the Convention will be reflected earlier, than in other matrices. Atmospheric monitoring (ambient air concentrations and atmospheric deposition) can thus be seen as one scientifically sound approach to help evaluate the Convention’s effectiveness (Gustin et al., 2016).
An overview of existing programmes and networks for monitoring atmospheric mercury, including standard operating procedures, and data gaps is available in the Supplementary Material to this guidance.

### 3.2.1. Mercury in air

As noted, the temporal and spatial scales of mercury transport in the atmosphere and its transfer to aquatic and terrestrial ecosystems depend primarily on its chemical and physical forms. There are three forms or species of mercury commonly found and measured in the atmosphere namely,

- Gaseous elemental mercury (GEM),
- Gaseous oxidized mercury (GOM)
- Particle-bound mercury (PBM)

with the last two species being operationally defined.\(^9\)

In addition to the above, atmospheric mercury can also be measured as “total gaseous mercury” (TGM) which equals to GEM and GOM combined.

Quantification of atmospheric deposition of the different forms of mercury to various underlying surfaces is useful in assessing the impacts on the environment and human health. Following emission, GEM can be transported long distances before oxidation and/or removal by particle and gas-phase dry deposition or scavenging by precipitation. Due to its relatively long residence time in the atmosphere, GEM can be transported and deposited to remote locations such as the Arctic (Skov et al., 2004; Sprovieri et al., 2005; Sprovieri et al., 2010) and Antarctic (Dommergue et al., 2010; Angot et al., 2016).

GOM and PBM have shorter atmospheric lifetimes than GEM and as a result are generally deposited closer to emission sources (Hedgecock et al. 2006; Jung et al. 2009). Oxidized mercury compounds often have a more local impact than elemental mercury because they are water-soluble, are more reactive and thus deposit more quickly (Hedgecock and Pirrone, 2004; Fu et al., 2015, Layman et al., 2020).

Measurements of mercury species such as oxidized and particle-bound mercury compounds are important as they help to improve the understanding of short-term oxidation processes regarding the removal of mercury from the atmosphere (Pirrone et al. 2008; Fu et al., 2015; Weiss-Penzias et al., 2016; De Simone et al. 2016).

Mercury speciation measurements are used in various networks, and there are standard operating procedures (SOPs) available and comparison studies between different monitoring networks have delivered satisfactory results (Steffen et al., 2012, Gustin et al., 2015; Sprovieri et al., 2016). Although they are very important in helping to understand the global mercury cycle as well as improving model output, speciation measurements are quite complex, costly, and require very skilled operators to avoid analytical interferences in GOM and PBM measurements (Gustin et al. 2015). Furthermore, recent information also demonstrates that GOM might be underestimated by this method (Gustin et al. 2020). Therefore, for the purpose of evaluating the effectiveness of the Minamata Convention, GOM and PBM are not considered as priority mercury species at this stage. There is a need for further scientific work on these methods to understand biases in existing methods and improve the comparability across measurement technique and therefore, will be more suitable under Tier 2 or Tier 3. It should be noted,

---

\(^9\) Operationally defined species: Their chemical and physical structure cannot be exactly identified by experimental methods but are instead characterised by their properties and capability to be collected by different sampling equipment (Schroender et al., 1998).
however, that several monitoring networks and research groups perform mercury speciation measurements, and are encouraged to continue doing so as their data will be helpful in answering questions listed in Table 2.1, and will provide valuable information to the Effectiveness Evaluation.

3.2.2. Atmospheric mercury deposition

After mercury is emitted into the atmosphere, it eventually returns to the Earth’s surface via wet and dry deposition processes. The pathway for wet deposition occurs when mercury is deposited with precipitation.

The amount of precipitation is the main driver of wet mercury deposition to aquatic and terrestrial habitats but, in areas with a lot of rainfall, the amount of deposition may be limited by the availability of atmospheric mercury (Prestbo et al., 2009; Weiss-Penzias et al. 2016; Sprovieri et al. 2017). GOM and PBM are water soluble and the primary atmospheric forms responsible for wet deposition of mercury in precipitation. Moreover, forest and land fires have an impact on the mercury release to the atmosphere and on wet deposition.

Wet deposition can be measured directly by collecting precipitation such as rain or snow and measuring the amount of mercury relative to the quantity of precipitation. Monitoring mercury in precipitation is an important way of determining inputs of mercury into aquatic and terrestrial ecosystems (Aas et al., 2019, Sprovieri et al., 2017). When compared to automated atmospheric mercury monitoring, wet deposition monitoring is relatively easy to start off with and is considered a very reliable method to achieve comparable data across different monitoring initiatives or programmes (Brown et al., 2010; Sheu et al., 2019). It should be noted that total mercury deposition is determined by both wet and dry deposition (Brown et al., 2010). However, with the appropriate equipment and analysis, wet deposition can be measured by a number of cooperators at many locations. This would allow wet deposition to be used for comparison for the Effectiveness Evaluation.

Dry deposition is the transfer of atmospheric mercury (either a gas or a particle) to vegetation, soil, water, and snow, controlled by the characteristics of the atmosphere, the surface, and the mercury species (Zhang et al., 2009). It is also possible to model dry deposition on the basis of RGM/GEM and GOM measurements using a number of methods, such as surrogate surface, litterfall, and throughfall measurements (Wright et al. 2016 and references therein). Measurements based on surrogate surfaces, micro-meteorological methods, and dynamic flux chambers are highly uncertain, while inferential approaches to estimate dry deposition and bi-directional air-surface exchange models are used, but is subject to uncertainties in measurements but also to errors due to underlying assumptions (Lyman et al. 2020; Zhang et al. 2009, 2016c, 2019). It should be noted that dry deposition is large during mercury depletion episodes in the Arctic, and it is dominating in dry arid climates (Steffen et al., 2015, Feng et al., 2020).

3.3. Tiered approach for atmospheric mercury monitoring

The tiered monitoring approach presents a framework to identify and prioritize monitoring needed to 1) determine whether mercury concentrations in air are changing over time, and 2) whether observed changes in concentrations may be attributable to controls on emissions, releases, supply and use of mercury effected by the Convention. This approach seeks to build-on/expand existing mercury air and wet deposition monitoring efforts to promote consistency in data collection and advance collaboration.
across sampling activities. The guidance offered in this chapter describes criteria to consider when deciding which measurement methods to use, the frequency of measurements, and where to potentially locate sites depending on the monitoring programme needs and objectives. For mercury in air monitoring the Tiered approach can be seen as follows:

**Tier 1** – The objectives of this tier are to estimate background and impacted levels and identify temporal trends. It documents trends and spatial distribution in air (TGM/GEM) and in wet deposition over broad geographic areas and provides information to inform atmospheric modelling (statistical and mechanistic). The measurement methods included in this tier are cost effective, practical, feasible, and sustainable. Tier 1 offers an entry point for Parties and relevant organizations who wish to pursue one or more of these sampling options (e.g., automated, manual, passive), where feasible, to contribute to the Effectiveness Evaluation.

**Tier 2** – The objective of Tier 2 includes source attribution, characterization of spatial patterns, and estimation of exposures and adverse effects. This tier explains temporal trends and attributes mercury sources to mercury concentrations in other matrices. Sampling is more intensive than Tier 1 sites and also relies more on ancillary measurements and models. Tier 2 will also enable "top-down" attributive analysis of TGM/GEM levels, using speciated mercury data and air-quality tracers. The more Tier 2 information are available, the more robust the Effectiveness Evaluation will become.

**Tier 3** – In Tier 3 the observational strategy is designed in such a way as to understand key processes affecting Hg fate and transport. This tier improves representativeness of the measurements and understanding of key processes (e.g., related to transformation and deposition) using new, advanced measurement techniques and sophisticated research. Where Tier 3 efforts and results are available, the information should be taken into consideration in the Effectiveness Evaluation.

### 3.4. Where should atmospheric mercury measurements be collected

Site characteristics can affect the concentration levels of mercury in air. Therefore, site selection is a critical part of any monitoring network’s design (Schmeltz et al., 2011) To support the Effectiveness Evaluation of the Convention, a variety of sites is needed. The selection of monitoring sites to support the Effectiveness Evaluation of the Minamata Convention, and help address policy questions 3 and 4 listed in the previous chapter, should be based on the sites’ potential to provide insights into changes in atmospheric mercury levels, to assess levels in sensitive ecosystems or vulnerable (Artic), and to help evaluate atmospheric models.

A diversity of site locations should be considered, as different types of sites may provide different types of information, including (a) background or remote, (b) rural, (c) urban and (d) contaminated or industrial sites. Background and remote sites, including forests not impacted by activities such as ASGM, provide information on determining long-term global trends and provide data for evaluating and refining transport models; rural sites provide information on mercury concentrations that are regionally representative as long as the influence of significant local pollution sources is limited; urban sites can provide information on non-point sources (pollution that comes from many places, all at once) such as Hg transported in cities, and will be useful to improve emissions inventories; and contaminated or industrial sites will assist with determining the effect on human exposure of communities living close to point sources (e.g. ASGM activities).
Mercury air and deposition monitoring site locations should prioritize collocation with other existing monitoring activities (for example, air quality sites and weather stations) to make use of available infrastructure and co-measurements with reactants important for Hg. Mercury air and deposition monitoring should also prioritize collocation with monitoring of Hg in other media to allow linkages to be quantified. Furthermore, in addition to spatial coverage, temporal coverage may also be a relevant element in the selection of monitoring sites so that areas with historic monitoring data, i.e. from a period before the implementation of the Convention, can be used as baseline.

One concept that has been enshrined in most proposed monitoring program structure is to have so-called “intensive” sites where detailed measurements are made and have these sites then coupled with a number of other sites (“cluster” sites) that can expand the data collection regionally and spatially. For atmospheric Hg, the intensive sites could include atmospheric mercury speciation measurements (gaseous Hg speciation, aerosols, and wet deposition), and therefore would need the infrastructure to support such measurements, while at the cluster sites only total gaseous Hg and weekly wet or bulk deposition could be measured. Given the development of passive sampling techniques, these could be used at the cluster sites allowing for the collection of data in remote areas where electricity is limited.

The number and location of the monitoring sites in any country should be determined based on the distribution of major urban areas, and take into account differences in the vegetation and should include coastal as well as inland sites (Evers et al., 2012). For example, Smeltz et al. (2011) suggested that the site locations be determined by the major ecoregions in the USA, as well as determined by other factors such as the likelihood for the site to experience change in concentrations over time. Additional criteria include their usefulness for model evaluation, and their importance in terms of evaluation of health risk to humans and wildlife (e.g. the Arctic). As noted above, intensive sites should also collect ancillary information besides Hg speciation, and should include existing Hg monitoring sites, where possible, such as those that are part of current national/regional networks or global mercury monitoring programs.

During the past two decades, a number of Hg monitoring sites have been established in Europe, North America and Asia as part of regional or global monitoring networks (Pirrone et al., 2003; Steffen et al., 2008; AMAP 2011, 2021; Fu et al., 2012; Sprovieri et al., 2016) using the information presented above as guidance when new sites were established. The need to establish a global network to assess likely southern-northern hemispheric gradients and long-term trends has long been considered a high priority for policy and scientific purposes in order to evaluate the impact of mercury pollution. Consistent globally distributed Hg observations will help understand trends in mercury concentration and deposition in different regions of the world. Building on and expanding on existing infrastructure will help improve the process of evaluating the effectiveness of the Convention. Examples of global, regional and national monitoring programmes are provided in the Supplementary Material to this guidance.

### 3.5. How to measure mercury in air: sampling and measurement methods

Several different methods are available for monitoring of air Hg. Selection of methods should be based on the purpose of monitoring. All methods employed in a monitoring program need to have been tested, intercompared and validated to ensure quality of data used for the Effectiveness Evaluation. This section aims to facilitate the selection of monitoring techniques that best meet monitoring needs and resource and logistical constraints, while recognizing the need for flexibility in the selection of techniques.
3.5.1. Active air sampling

Active air sampling methods involve ambient air pulled through a pump at a constant flow rate through an active material, otherwise known as a trap.\(^{10}\) The active material contained in these traps is often gold, but other materials such as sand mixed with gold or carbon are also used. Once the sample has been collected on the active material for a set amount of time, the mercury adsorbed is removed using thermal desorption and spectroscopic detection.

Active sampling of this kind can be undertaken in an instrument that both collects the sample and performs the analysis in situ (automatically) or can be undertaken by collecting the sample actively and then performing the analysis at a separate location or laboratory. The differences in the methods are explained below as (a) Automated Mercury Air Measurements; and (b) Manual Mercury Air Measurements.

(a) Automated mercury air measurements

Currently, automated mercury measurements, which can be used for GEM/TGM and speciation measurements (see section 3.2.1), are recorded with different commercial instruments available from various manufacturers that are capable of detecting mercury at very low concentrations in air, i.e. micrograms of gaseous pollutant per cubic meter of ambient air. These instruments have high temporal resolution, low limits of detection, established and proven quality assurance and quality control protocols (Brown et al., 2010; Angot et al., 2014; Gustin et al., 2015; Sprovieri et al., 2016; Slemr et al., 2020).

The following spectroscopic detection techniques are most commonly used for automated monitoring of mercury in air, either as total gaseous mercury (GEM + GOM) or as gaseous elemental mercury (GEM) are:

- Cold Vapour Atomic Absorption Spectroscopy (CV-AAS)
- Cold Vapour Atomic Fluorescence Spectroscopy (CV-AFS)

Mercury is present in very low concentrations in air, i.e. it is present as nanograms of mercury per volume of air (ng/m\(^3\)) at standard pressure and temperature, and it has a particular strong absorption/emission line at 253.7 nanometer (nm). In most cases and with the exception of Zeeman AAS and laser techniques, the sensitivity for AAS and AFS is not sufficient for direct measurements of ambient concentrations and, therefore, mercury is pre-concentrated on gold coated surfaces (Gustin et al., 2010; Amos et al., 2012; Gustin et al., 2015). Both techniques are however sensitive to molecular species (e.g. ozone, sulphur dioxide, organics) that absorbs in the ultraviolet (UV) range close to 253.7 nm. Unlike AAS, AFS is less influenced by these molecular interferences and does not require any sort of correction scheme. AFS is more sensitive in comparison to AAS, but it requires pure Argon (Ar) or Helium (He) gas during the desorption and detection step, whereas AAS uses mercury free air or nitrogen instead (Gustin et al., 2015).

The pure gold trap method with AFS has been chosen by national and international networks for use at background and regional sites and is cited extensively in the scientific literature (Sprovieri et al. 2016; Martin et al., 2017, Xu et al., 2017). At the same time, for many other applications where low detection limits are not required, less concern about interference and a lower QA threshold is set, the use of

\(^{10}\) The Zeeman-AAS method does not use a trap.
different instrument methods are good choices for monitoring at artisanal gold mining and industrial locations, etc. Automated instruments are used widely within different monitoring networks and programs and can generate data that will be comparable (Steffen et al., 2012, Sprovieri et al., 2016). The cost (investment and running costs) associated to the use of automated analysers are substantially higher compared to other methods (passive samplers, wet deposition, and manual method). However, these instruments are able to deliver high frequency data in a short time span from as little as 5 seconds to 5 minutes.

(b) Manual mercury air measurements

With this technique, mercury in the atmosphere is collected manually on an adsorbent material over 24-hour periods, or weekly, at a constant flow rate using a pump. The sampling instruments normally use a gold tube scavenger and small air pump, and it is portable. For this reason, this technique is less site restricted. This method is relatively easy to setup and operate but requires a covering or shield to protect the trap from contamination or interference by weather elements.

Analysis of TGM in ambient air is possible when the sample is analysed after exposure using thermal desorption and spectroscopic detection in a laboratory. Additionally, since the scavenger material, which is generally used in manual active sampling (e.g. gold cartridge), catches almost all amount of mercury in air, the manual active sampling has very little or almost no isotope effect. Thus, collected air sample by manual active method is useful for isotope ratio analysis.

The measurement principle applied for manual active method is same as that for automated active method and interferences are caused by same environmental factors (e.g. by ozone and high humidity). The data from this technique can be calculated from a calibration curve based on measurements of mercury gas that are accurately collected from a saturated mercury gas generator (bell-ger). Minimum time resolution is dependent on the measurement instrument used (around 3 hours for CVAAS and 1 hour for CVAFS depending on analytical condition and concentration of the air) and, in most cases, within the resolution of automated active and passive sampling methods. Unlike automated sampling, manual active sampling method is analysed by using an analytical instrument which is installed in a laboratory separately from sampling equipment. Analytical instruments are costly, but this method is cost-effective because scavenger materials are low cost and analytical instruments can be shared while being able to place sampling equipment in multiple sites.

The temporal resolution of this method is often lower than that for automated mercury monitoring, but 24h-average is suitable for long-term trend analyses. It should be noted that achieving quality results with this method, would require consistent low blanks and an operator with trace-clean technique experience. Interference from ozone and high humidity can also influence the performance. QA/QC procedures such as field blank and co-located sampling for instruments and staff proficiency are important because multiple equipment and personnel are involved. General approaches for these QA/QC procedures are well documented ( Munthe et al., 2001; Brown et al., 2010; MOEJ, 2011), and methodological comparisons for automated active and passive sampling are in progress (e.g. in Japan).

3.5.2. Passive sampling

The development of methods for passive air samplers (PAS) for gaseous mercury has increased recently. While it is not possible to produce data at the same temporal resolution as automated or manual instruments (McLagan et al., 2016), PAS have been shown that they can produce air concentrations of
mercury accurately and comparably with active air monitoring methods. PAS can increase spatial resolution of air concentration data and contribute to Hg source characterization. PAS work by diffusive uptake through a diffusive surface and accumulation of gaseous Hg onto an adsorbent scaffolding. The peculiarity of the passive samplers relies on the unassisted molecular diffusion of gaseous agents following Fick’s First Law (i.e. volatile vapours of elemental mercury). Unlike actively pumped sampling, passive samplers require no electricity, have no moving parts, no pump operation or calibration, and are simple to use and low cost (CNR, 2019). PAS can be deployed at background, remote, urban, hotspots sites and without worry about media failure as a new sampler with active material is used each time. After exposure, the PAS can be analysed with well-documented and credible methods in any analytical laboratory (Wängberg et al., 2016, McLagan et al., 2016, 2017, 2018 Macagnano et al., 2018). It should be noted that depending on the exposure period chosen, PAS have lower temporal resolution than other methods and its sampling rate can be affected by wind speed and temperature, which must be factored into the calculations (Mclagan et al., 2017). Quality control of the samplers is necessary even when a new product is used (i.e. it cannot be left to providers), and appropriate QA/QC process is important for the entire survey and analysis, including at laboratories. Laboratory measurements of passive air samples are not as easy as other types of samples and interlaboratory comparisons may not be possible. However, a recent study shows good insight into how to undertake intercomparisons of data generated with different PAS (Naccarato et al., 2021). Moreover, PAS have shown to be useful for low level monitoring as well as at high concentration sites or hotspots, and they can assist with understanding contaminated sites emissions to air, specifically describing concentration gradients. For initiatives that have no mercury air monitoring program or previous mercury air experience, passive sampling is considered a suitable method to start with (Tier 1; see tiered approach in chapter 2 and annex) or to complement to a limited number of active analysers. In the context of the Effectiveness Evaluation of the Convention, if deployed at the relevant sites, PAS will contribute to answering questions related to spatial variability, trends and emissions and impacts on local ASGM communities.

3.5.3. Wet deposition sampling

Mercury wet deposition sampling can be carried out using a variety of commercially available precipitation collectors for either wet only or bulk collection (Prestbo et al., 2009, Brown et al., 2010; Gichuki et al., 2013, Brunke et al., 2016, Weiss-Penzias 2016; Risch et al 2017.). Wet only samples are collected using Teflon or borosilicate glass bottles. When operating a wet only sampler, the precipitation sensor is activated during a precipitation event, and the lid moves off the funnels and comes to rest. When an event stops, the sensor dries out, and the lid returns to cover the funnels. For bulk collection, the sampler is open to the atmosphere constantly (Brunke et al., 2016; Sheu et al., 2019). The wet-only samplers have the advantage of avoiding particle dry deposition although the contribution to the measured wet deposition fluxes from gaseous or particulate mercury species is probably not large in non-industrialised or non-urban areas. For extended sampling periods it is also necessary to prevent significant gas phase diffusion of Hg$^0$ to the surface of the collected sample where it could contribute to the mercury content of the sample via oxidation to water-soluble forms. This can be easily done using a capillary tube between the funnel and the bottle. Shielding of the sample bottles from light is also necessary to avoid photo-induced reduction of the mercury in the precipitation sample. Wet deposition sampling collects all water-soluble atmospheric mercury species (Hg$^0$ + MeHg). The amount of Hg
depositing from the atmosphere to the ecosystems provide a useful proxy for impacts of Hg emissions (Selin 2018; Travnikov et al., 2017).

When starting with wet deposition monitoring, the following factors should be considered when choosing a sampling location: (a) availability of stable electricity when using a wet-only collector (wet-only collectors can run on solar and battery power in some locations, without access to power lines; bulk collectors require no electricity), (b) access to laboratory facilities to prepare samples, including treating glassware with acids and other chemicals, and to analyze the collected samples, (c) access to proper shipping, including portable sample trays and coolers for collecting and transporting field samples, (d) skilled operator to conduct analysis, (e) availability of meteorological data at sampling site, (f) fridge or other storage facility for samples, (g) site selection so it is following the general recommendations for precipitation measurements from WMO.

3.5.4. Dry deposition sampling

Direct measurement of mercury dry deposition is technically challenging but can be done by micro-meteorological methods (Brooks et al., 2006, Skov et al., 2006). However, it is also possible to model dry deposition on the basis of RGM/GEM and GOM measurements (see section 3.2.2). Although no methods currently exist to measure mercury dry deposition in a network configuration, dry deposition measurements may be useful as part of monitoring activities in Tier 2 and Tier 3 (see annex). Further expert consultation is necessary to consider if and how dry deposition can contribute to the Effectiveness Evaluation.

3.6. Frequency and duration of sampling

The following sampling and exposure times are guidelines for the various sampling methods and sites locations depending on how the data will be used.

(a) Automated measurement with CV-AFS/AAS

- A sampling time of 5 min to 15 min where the average concentration is lower than 1.0 ng/m³ (mostly sites located in the southern hemisphere).
- Half-hourly averages of quality-controlled data can be used for process studies and hourly averages for trend analyses.

(b) Manual measurements

- A sampling time of 24 hours and a constant flow rate is sufficient as several networks are currently using this parameter and available data can be used as comparison.
- Frequency should be determined appropriately (weekly, monthly, etc.) so that comparable data to automated sampling are obtained, based on local conditions.

(c) Passive sampling

- To generate data with PAS, an exposure time of 1 to 3-months is suitable for the Effectiveness Evaluation for a global network (McLagan et al., 2016, 2018).
If sufficient resources and manpower are available, the sampling time can be increased (e.g. to monthly exposure) which will increase data points and be very helpful in the Effectiveness Evaluation process.

(d) *Wet deposition*

- Weekly sampling is recommended to provide an integrated 24-hour 7-day sample as this is the preferred procedure currently in use within most wet deposition networks.
- What is of importance is that once an exposure period has been decided on, the site operators should keep to that schedule.

3.7. Quality assurance and quality control for field air monitoring operations

(a) *Instrumentation*

It is important to ensure that the correct installation and operation of automated/manual passive air samplers instruments and wet deposition collectors are followed. The manufacturers’ guidelines or relevant SOPs within networks should be followed. Instruments must meet minimum requirements for e.g., sensor sensitivity to chemical inertness.

(b) *Sample collection and handling*

Specific quality control procedures that prevent contamination from occurring during sample collection and handling for the various monitoring techniques include:

- Wearing disposable plastic gloves whenever handling precipitation collectors, passive samplers and transferring samples from field sites;
- Properly transporting samples by “double bagging” samples after collection;
- Checking for, and documenting, sample leaks in the field, during shipping, and upon receipt at the laboratory.

(c) *Field notes*

Field notes describing the sampling method and overall conditions, such as any deviations from the standard sampling method and meteorological and other environmental conditions that may affect the measurements, should be written down and kept safe and dry with the samples, with copies kept at separate location.

(d) *Sample storage and shipping*

Proper storage and shipping methods must be used to preserve the chemical and physical integrity of samples. Quality control procedures for this purpose include:

- For wet deposition samples, maintaining samples (precipitation collected) in cooled containers while in transit and when stored in laboratory;
- Weighing wet deposition samples to determine sample volume at the station and at the laboratory in order to detect leaks in transit;
- Precipitation samplers should not be stored longer than 6 months before it needs to be analysed, even if they are properly stored and preserved;
Passive samples should be stored in a cool dry place after collected from the field in double bags and sealed tightly.

(e) Blanks

Field blanks are to be collected on a regular basis to ensure that sampling methods and materials do not interfere with sample chemistry. It is recommended that blanks be collected randomly at every site. For manual active sampling monitoring methods, field blank test is performed regularly (for example, once every 10 times). For wet deposition monitoring the blanks are to be collected by pouring an aliquot of deionised water into a dry sample container (e.g., bucket, bag, funnel-and-bottle) for a sampling period during which no precipitation occurred. The aliquot should be submitted to the laboratory in the same manner as precipitation samples. For passive samplers, the samples may be exposed to other samples, but they must be hermetically sealed and isolated from atmospheric air.

3.8. Ancillary data

Ancillary data are collected to allow the (mercury) data to be understood in a valid manner; they are not indispensable for using the data but serve as additional information valuable for interpreting it. The most relevant ancillary data for mercury in air monitoring are: (1) meteorological variables such as temperature, pressure, relative humidity, wind direction and wind speed and (2) chemical variables such as carbon monoxide (biomass burning), sulphur dioxide (volcanic activity) or ozone (Arctic Mercury Depletion Events), which can be used to identify sources and atmospheric processes. For the collection of ancillary data, the following WMO GAW Guidelines will be useful (GAW Report 183 WMO-2008, GAW Report 192, WMO-2010, GAW Report 201, WMO-2014, GAW Report 204, WMO-2012). The ancillary data should be collected at the same sites and stored with the same metadata and data format as the mercury data.

3.8. Management, analysis and evaluation of atmospheric mercury data

The following tools will provide Parties and relevant organisations with a more holistic picture of the state of mercury in air by adding value to the monitoring data that is collected.

3.8.1. Local, regional and hemispheric trend analysis

- Atmospheric concentrations and wet deposition data.
- Mann-Kendall non-seasonal trend analysis, preferentially pre-whitened, or machine learning methods (e.g. Empirical Wavelet Methods)
- For trend analysis, it could make sense to look at groups of stations for a region, latitudes or even a complete hemisphere.
- Depending on how the data is collected, the minimum number of years and minimum data coverage to calculate trends is recommended to be 5 years or more and minimum monthly coverage of 60% (with the values giving an idea of what is deem sufficient when trends are reported in literature)

3.8.2. Data based analysis

- e.g. PMF (Probability Mass Function) analysis of sources using other measured species indicative of major Hg sources (e.g. SO2, CO2).
• Based on statistical methods this approach uses the secondary measurements to identify different sources. e.g. high Hg and high SO\textsubscript{2} could be a volcanic or coal combustion source. High Hg and high CO could be a burning source.

• The PMF method could be used once a single year with data coverage of (>70\%) is achieved as a means of source/sink appointment.

3.8.3. Source receptor relationship based on footprints and trajectory analysis

• Analysis of source regions using backward modeling datasets generated by Lagrangian models like HYSPLIT or FLEXPART. Backward trajectories or 3-dimensional footprints of air parcels released at the measurement location and followed up to 5-20 days backwards in time.

• By combination of measurements and information on air mass origin it is possible to determine source/sink regions based on a PSCF (Potential Source Contribution Function). One can also combine several stations for a comprehensive map. The next step would be a feasibility analysis to explore the potential for gaining more quantitative emission estimates by inverse algorithms.

3.9. Conclusions

This chapter has identified different methods Parties and organizations can use to monitor atmospheric mercury and generate comparable data to support the Effectiveness Evaluation of the Convention. Monitoring of atmospheric mercury has been ongoing for decades but not all regions are equally covered with the biggest data gaps in the southern hemisphere. A tiered approach is proposed and gives Parties and organizations an opportunity to start, expand or improve their monitoring programmes in such a manner that comparable data can be generated to support the Effectiveness Evaluation. Moreover, the tiered approach also breaks down the monitoring requirements in such a manner that new atmospheric mercury monitoring initiatives have an opportunity of joining one of the several existing monitoring programmes and networks, thus drawing from the experience and information at hand that these established networks can provide.

Automated atmospheric mercury measurement is the preferred method within existing monitoring networks. While the instruments used in automated measurements are capable of detecting very low concentrations of mercury, these instruments are expensive and alternative options are available that can also deliver comparable data. Passive and manual sampling of atmospheric mercury are two such options, even if at a lower temporal resolution as compared to automated systems.

Depending on the specific needs of the monitoring initiative, this guidance puts forward different methods at Tier 1, as the minimum step to start generating comparable atmospheric mercury data. Wet deposition of Hg from the atmosphere is one of the methods included at Tier 1 level. The method is reasonably well understood and sufficient results have been achieved in networks, as well as on a global scale, through various studies and intercomparison exercises. Therefore, scientifically sound, and cost-effective methods and techniques to determine mercury concentrations in air are available and can generate comparable data.

Another important factor to take into account when performing mercury air monitoring is the location(s) where monitoring will take place. Monitoring at a variety of sites will provide a more comprehensive picture of the levels of Hg in the atmosphere. It is therefore important for each
monitoring initiative to identify sites that can provide insights into changes in atmospheric mercury levels over time, including relevant and sensitive ecosystems. Carefully selected sites can also help develop more robust atmospheric models and fill data gaps.

Beyond sampling and analysis, for any monitoring programme to be successful, a strong quality assurance and quality control (QA/QC) program is needed. A wealth of experience on key elements and processes related to QA/QC is available from existing atmospheric mercury monitoring programmes and networks as seen in this chapter and Supplementary Material.

The elements put forward in this guidance will help answer various operational questions listed in Chapter 2 with regard to atmospheric mercury monitoring. Furthermore, this guidance provides Parties and relevant organizations with the means of starting, improving or expanding on their initiatives for monitoring atmospheric mercury that will enable them to deliver comparable data that can support the Effectiveness Evaluation of the Minamata Convention.
Chapter 4. Biota Mercury Monitoring

4.1. Introduction

Mercury emitted to the air and released to water and land can be retained in the environment for years to millennia and may be transported across great distances, where its fate is complex as it moves through and across terrestrial and aquatic ecosystems (Driscoll et al. 2013; Kocman et al. 2017).

Inorganic mercury from natural or anthropogenic sources becomes more toxic in the environment when it is converted to methylmercury (MeHg) by microbes.

Methylmercury readily biomagnifies through both aquatic and terrestrial ecosystems, resulting in increasing concentrations as it moves from the base of the food web to higher trophic levels (Eagles-Smith et al. 2016b; Figure 4.1). Generally, each trophic change in the food web accounts for roughly an order of magnitude (10x) of increase in MeHg concentrations, with the largest enrichment step occurring between water and plankton in aquatic systems (Lee and Fisher, 2016). As a result, meso and top predators in a food web, such as fish, reptiles, birds, and mammals, may have MeHg concentrations in their tissues that are many orders of magnitude higher than the concentrations found in the surrounding environment (often > 10⁶ to 10⁷ higher).

Studies show that the biomagnification (trophic level exchange) and bioaccumulation (body burden accumulation over time) of MeHg adversely affect the reproductive success of many wildlife species, impacting multiple taxa across many habitats and geographic areas of the world (e.g., birds in North America; Ackerman et al. 2016, Evers 2018, Dietz et al. 2019). Moreover, dietary uptake of methylmercury by humans, primarily through the consumption of fish, but also of marine mammals and birds, is a primary health concern (Transande et al. 2016; Dietz et al. 2018; Fielding et al. 2021).

Monitoring of mercury in biota can inform policy-making at various levels and across sectors. It can also help support the objectives of the Effectiveness Evaluation of the Minamata Convention through the information categories identified in chapter 2 (see Table 2.1).

Ongoing biomonitoring programs provide valuable information for the Effectiveness Evaluation and the breadth of existing data will provide a basis for establishing comparable bioindicators, geographic areas of interest and a baseline for estimating change, as well as for identifying data gaps. However, there are some challenges in using existing data that were not necessarily designed for describing spatial patterns or standardized tracking of temporal trends or linking with anthropogenic mercury sources, even if they can be alleviated by introducing standardized tracking over time and by adding suitable ancillary measurements. The extent to which the existing information may be used in the Effectiveness Evaluation will depend on the degree of flexibility in the use of existing biotic mercury concentrations as baselines – that may be an important determinant for establishing time series and developing spatial patterns. Continuous monitoring programs that are designed to produce long times series with comparable methods and ancillary data will be particularly valuable in this effort. Furthermore, linkages of biotic Hg concentrations to anthropogenic sources and uses of mercury as identified in the Minamata Convention can be conducted with statistical analysis and modelling, given the availability of suitable ancillary data that can be linked to different source types (Harris et al. 2007; Knightes et al. 2009; Dietz et al. 2019; Schartup et al. 2019).

Fostering international collaboration and coordination among national projects will be crucial to create harmonized regional approaches and to strive, where possible, to integrate biomonitoring activities in
an interdisciplinary manner (i.e., including air and human biomonitoring as well as biota) to assess ecological and human health risk that can be merged to illustrate regional and eventually global spatio-temporal patterns.

This chapter provides a brief overview of our state of knowledge with regards to existing data in fish and wildlife, and proposes a way forward by which biomonitoring data can be used to support the Effectiveness Evaluation. It offers scientific and technical considerations for the selection of bioindicators and monitoring sites, as well as on ecosystem sensitivity and its importance to assessing threats to the environment and human health and the need for ancillary measurements. Finally, this chapter explores a possible tiered approach to monitoring mercury in biota with a view to supporting the Effectiveness Evaluation of the Convention.

![Figure 4.1. Examples of trophic level exchange for freshwater and marine ecosystems.](image)

### 4.2. State of knowledge

Mercury exposure has been well documented in fish and wildlife around the world. Published mercury concentration data for the target biota of the Minamata Convention exceed 530,000 data points and represent the world’s oceans and continents. Biotic mercury concentrations are most robust for fish, for both marine and freshwater ecosystems, and the number of analysed samples are known to be much greater when including unpublished governmental and other datasets.

Numerous studies, particularly recent ones, document adverse impacts across many fish species. In fish, adverse impacts of MeHg exposure include reproductive behavioural, and immunological impairment (Depew et al. 2012a, Scheuhammer et al. 2015; Carvan et al. 2017). Elevated methylmercury concentrations may impact recreational and commercial fisheries by reducing the viability and sustainability of fish populations, especially those in ecosystems with high sensitivity to Hg methylation (Evers et al. 2007) and at higher trophic levels – due to biomagnification effects.
In birds, numerous studies document reduced reproductive success, behavioural change (e.g., reduced time incubating), and neurological problems (e.g., ataxia) (Depew et al. 2012a,b; Ackerman et al. 2016, Whitney and Cristol 2017, Evers 2018; Cristol and Evers 2020).

In mammals, elevated MeHg concentrations can result in biochemical changes in the brain, ataxia, and reduced reproductive output (Ronald et al. 1977; Dietz et al. 2013, Evers 2018, AMAP 2021). The effect thresholds for marine mammals are poorly understood, but based on mercury effect thresholds for terrestrial mammals, there could be significant adverse impacts on the reproductive success of marine mammals (Dietz et al. 2019).

Existing mercury biomonitoring networks for biota that have ongoing and standardized measurements that can be used for objectives such as tracking temporal trends are relatively rare as documented by a review by UNEP (UNEP 2016). An overview of existing biota monitoring programmes, networks and databases are provided in the Supplementary Material.

4.3. Proposed elements for monitoring mercury in biota

The development of a sustainable and long-term global biomonitoring initiative linking existing data on mercury levels in biota to the Effectiveness Evaluation of the Minamata Convention could focus on (i) identifying areas that have regional data gaps so new programs can generate data that meet statistical power for confidence in understanding spatial patterns, including those that incorporate ecosystem sensitivity (Evers et al., 2011b) and temporal trends (Bignert et al., 2004; AMAP 2011) and (ii) expanding existing monitoring programs to support trend analysis and modelling, and establish linkages to Hg source types.

To this end, an understanding of (i) key environmental attributes for mercury transport, methylation, and bioaccumulation, (ii) exposure pathways and effects of mercury on target bioindicators, and (iii) contributing factors that influence mercury transformation, bioaccumulation and biomagnification will help normalize observed Hg concentrations. Biota monitoring activities can contribute to answering the operational questions identified in chapter 2.

4.3.1. Selection of bioindicators

Mercury monitoring using biotic media requires the careful selection of aquatic and terrestrial bioindicators and associated tissues that can realistically respond to key objectives of identifying temporal trends, spatial patterns, and linking with mercury source types.

A key initial step in bioindicator selection is to decide whether the aim of biota monitoring is linked to ecological health endpoints and/or human exposure assessments. It is often possible to select organisms that provide monitoring data for both purposes. Careful selection of bioindicators will further provide information about the potential impacts of MeHg contamination on biodiversity, including threatened species (e.g., IUCN Red List of Threatened Species™),11 key species, as well as species of national interests for their conservation and protection. Monitoring of such species will also provide insights on the relationship between mercury and additional stressors such as habitat degradation, climate change, and overharvesting. For example, ASGM activities in tropical systems significantly contribute to environmental methylmercury loads as well as severely altering habitat quality in areas with high endemism (Sayers et al., in preparation) – such locations are therefore of interest for biomonitoring.

11 https://www.iucnredlist.org
The taxa of interest for the Minamata Convention, based on Article 19 (b), include the “modelling and geographically representative monitoring of levels of mercury and mercury compounds in vulnerable populations and in environmental media, including biotic media such as fish, marine mammals, sea turtles and birds, as well as collaboration in the collection and exchange of relevant and appropriate samples”. The extensive data on Hg in biota found in the published literature, can inform the selection of bioindicators for monitoring. Informed selection can ensure cost-effective comparability at regional and global scales.

Table 4.1 lists a number of species and species’ groups that are well described and may serve as useful indicators for ecosystem health and human exposure assessment, categorized within their respective biomes and associated aquatic ecosystems (Evers et al. 2016). Appropriate tissue types for varying objectives are shown in a tiered approach in Annex 1. Biomonitoring for tracking temporal trends should be consistent with species, tissue, and location sampled, sampling methodologies, and analytical approaches.

**Table 4.1.** Examples of trophic levels 3 and 4 biota that could serve as bioindicators with major biomes and associated nearshore areas (based on Evers et al. 2016).*

<table>
<thead>
<tr>
<th>Terrestrial Biomes and Associated Marine Areas</th>
<th>Ecological Health Bioindicators</th>
<th>Ecological Health Bioindicators relevant for assessment of potential human exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freshwater Birds</td>
<td>Marine Birds</td>
</tr>
<tr>
<td>Arctic Tundra and Arctic Ocean</td>
<td>Loons, Songbirds</td>
<td>Fulmars, Murres</td>
</tr>
<tr>
<td>Boreal Forest-Taiga and North Pacific and Atlantic Ocean</td>
<td>Loons, Eagles, Osprey, Songbirds</td>
<td>Osprey, Petrels</td>
</tr>
<tr>
<td>Temperate Mixed Forest and Pacific and Atlantic Ocean</td>
<td>Loons, Egrets, Herons, Eagles, Osprey, Terns, Songbirds</td>
<td>Osprey, Terns</td>
</tr>
<tr>
<td>Tropical Rainforest and South Pacific and Atlantic and Indian Ocean</td>
<td>Egrets, Herons, Kingfishers, Songbirds</td>
<td>Albatrosses, Frigatebirds, Shearwaters, Terns, Tropicbirds</td>
</tr>
</tbody>
</table>

* Note: trophic level 3 or 4 young individual fish (<2 years) can be used for the objective of tracking temporal trends.
An important step towards developing comparable biotic monitoring data to inform the Effectiveness Evaluation of the Convention is to define regional biological species for monitoring in order to minimize the effects of species-specific physiological differences. For example, there are several game fish species that are found in northern Europe and North America that accumulate significant amounts of Hg due to their high trophic level and are frequently used by Hg biomonitoring programs (DePew et al. 2013; Eagles-Smith et al. 2016; Olk et al. 2016). To be able to potentially explain the main drivers behind the spatial patterns and temporal trends of fish Hg concentrations, and how these patterns and trends change under influence of different and emerging drivers (including environmental / climate change and deposition change in addition to changes in emissions and releases), a set of minimum target information could be developed. While adult game fish may be useful for characterizing spatial gradients and estimating exposure and adverse impacts for human exposure and ecohealth, younger fish are best for tracking temporal trends to avoid confounding effects due to shifts in diet and trophic structure.

Based on the knowledge of existing biotic Hg data and only using comparable data (e.g., trophic level 3 or 4 species that can be regularly sampled for comparable purposes for understanding spatiotemporal patterns) for relevant terrestrial biomes and associated marine areas, a matrix of available data can respond to questions related to spatial patterns, temporal trends, and linkages with mercury source types. While monitoring mercury in trophic level 3 and 4 biota can be useful for assessing potential mercury exposure for humans and top predators, which play important roles in maintaining ecosystem health and high levels of biodiversity (Sergio et al. 2005), understanding temporal trends is more complex. Rich supporting information on the processes that affect bioaccumulation and mercury availability is therefore needed to conduct robust trend analysis and separate anthropogenic and androgenic influences from each other. Such information is readily available from several long-term programs in the Northern Hemisphere and they have already been used for trend assessment on a pan-regional scale (AMAP 2021). When establishing new programs primarily intended for analysing trends, detecting short-term (<10 years) trends in changes of mercury in biota are best viewed through young individuals (<2 years of age) where age and therefore bioaccumulation (important for high risk, long-lived species that can increase Hg body burdens over their lifespan) is not such a significant confounding factor. Trophic level 3 and 4 fish may still be used for this objective, but younger rather than older individuals could be sampled. The use of trophic level 3 and 4 fish, both young (for temporal trends) and adults (for spatial patterns) can simplify sampling efforts.

Final selections of target biota for monitoring Hg and its impacts on the environment and human health should be evaluated for their life history characteristics, as well as their plasticity in foraging ecology and habitat, spatial use and movement patterns, variability in growth rates, temperature and general water quality tolerances and geographic distribution.

4.3.2. Tissue types

The selection of tissue types varies according to the study objectives, taxon being monitored, and characteristics of the monitoring sites. Examples of the proper selection of tissue type are well-established with associated information about the percent methylmercury content in the tissue and the preferred type of tissue preparation (Table 4.2). Most muscle, blood, egg and keratin-based (e.g., scutes, feathers, and fur) tissues primarily contain methylmercury. This is important for the simpler and more cost-effective laboratory analyses of total mercury concentrations that can assume 95% or more
methylmercury content (with some exceptions such as whole body analyses of small fish). Field protocols are available for all tissue types (see Supplementary Material).

Table 4.2. Major biota groupings and tissues recommended for MeHg monitoring. Except for the whole body samples, all tissues can be non-lethally sampled.

<table>
<thead>
<tr>
<th>Biota Group</th>
<th>Tissue Type</th>
<th>% MeHg</th>
<th>Sample preparation type*</th>
<th>Analysis type</th>
<th>Source reference for % MeHg</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish</td>
<td>Muscle fillet</td>
<td>&gt;95% (but varies on average as low as 65%)</td>
<td>ww or dw</td>
<td>THg</td>
<td>Bloom 1992; Lescord et al 2018</td>
<td>Recent evidence indicates that %MeHg may be lower for some fish species and for some cooking approaches (Wang et al. 2013) so to confirm the expected amounts. 10% of fish should be analysed for MeHg content.</td>
</tr>
<tr>
<td></td>
<td>Muscle biopsy</td>
<td>&gt;95% (but varies)</td>
<td>dw</td>
<td>THg</td>
<td>Peterson et al. 2004</td>
<td>dw is best because of moisture loss concerns. Muscle biopsy to muscle fillet has a $r^2 = 0.96$. Biopsy plug depth may impact Hg measured – 5mm plugs are best below dorsal fin (Cizdziel et al. 2002) and are without skin and adipose tissue.</td>
</tr>
<tr>
<td></td>
<td>Fin clips, muscle fillet and whole body</td>
<td>varies</td>
<td>dw</td>
<td>THg or MeHg</td>
<td>Cerveny et al. 2016</td>
<td>There is a significant correlation between fin clips and muscle fillet/whole body ($p&lt;0.01$).</td>
</tr>
<tr>
<td></td>
<td>Blood</td>
<td>&gt;95%</td>
<td>ww or dw</td>
<td>THg</td>
<td></td>
<td>Assumed to be &gt;95% MeHg based on other vertebrates.</td>
</tr>
<tr>
<td>Sea Turtles</td>
<td>Scutes / Carapace fragments</td>
<td>~10%</td>
<td>fw (or dw if scutes need washing)</td>
<td>THg</td>
<td>Rodriguez et al. 2019</td>
<td>While scutes are keratinized material the %MeHg may be relatively low and needs more data.</td>
</tr>
<tr>
<td></td>
<td>Blood</td>
<td>&gt;95%?</td>
<td>ww or dw</td>
<td>THg</td>
<td></td>
<td>Assumed to be &gt;95% MeHg based on other vertebrates.</td>
</tr>
<tr>
<td></td>
<td>Muscle</td>
<td>&gt;95%?</td>
<td>ww or dw</td>
<td>THg</td>
<td></td>
<td>Assumed to be &gt;95% MeHg based on other vertebrates.</td>
</tr>
<tr>
<td>Birds</td>
<td>Blood</td>
<td>&gt;95%</td>
<td>ww or dw</td>
<td>THg</td>
<td>Rimmer et al. 2005; Edmonds et al. 2010</td>
<td>Elimination of MeHg in blood comprises an initial fast phase, with half-time of 1 day, and a slow terminal phase with half-time between 44-65 days. Molt is a crucial factor in determining the rate of MeHg elimination (Monteiro and Furness 2001).</td>
</tr>
<tr>
<td></td>
<td>Feather</td>
<td>~100%</td>
<td>dw or fw, whole feathers or</td>
<td>THg</td>
<td>Burger 1993</td>
<td>Use feathers with caution; see Peterson et al. (2019) for a tool and guidelines for feather.</td>
</tr>
<tr>
<td>Sample Type</td>
<td>Purity (%)</td>
<td>Form</td>
<td>THg</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eggs</td>
<td>&gt;96%</td>
<td>dw or ww</td>
<td>THg</td>
<td>Ackerman et al. 2013 (96% for 22 species) ww and dw can be problematic if eggs are not collected immediately after laying (Dolgova et al. 2018).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eggshells and membranes</td>
<td>&gt;95%</td>
<td>dw</td>
<td>THg</td>
<td>Membranes are assumed to be primarily MeHg, but shells are entirely inorganic Hg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle</td>
<td>&gt;95%</td>
<td>ww or dw</td>
<td>THg</td>
<td>MeHg comprised over 99% of total Hg in breast muscle of waterfowl (Sullivan and Kopec 2018).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>&gt;90%</td>
<td>dw</td>
<td>THg</td>
<td>Wageman et al. 1998 Muktuk (in marine mammals) includes layers of skin and blubber</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fur or hair</td>
<td>&gt;90%</td>
<td>dw (or fw if fur is not washed)</td>
<td>THg</td>
<td>Evans et al. 2000 Use fur with caution; fur/hair may not relate to blood and muscle depending on growth patterns (Peterson et al. 2016a).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle</td>
<td>&gt;90%</td>
<td>ww or dw</td>
<td>THg</td>
<td>Wagemann et al. 1998</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Reported as wet weight (ww), dry weight (dw), fresh weight (fw), or fresh wet weight (fww) analyses. Fw denotes keratin-based samples that are not cleaned or dried prior to total Hg analyses.

### 4.3.3. Monitoring sites

Certain ecosystem conditions (primarily those with an aquatic component, especially wetlands) can encourage the production and bioavailability of MeHg in the environment. Bacteria often produce more MeHg when moderate amounts of sulphate and low oxygen (hypoxic or anoxic) conditions are present to provide optimal conditions for the metabolic processes of the bacteria (Hsu-Kim et al. 2013, 2018).

Environmental factors such as pH, dissolved organic carbon, total suspended solids, and sulphur concentrations are important in influencing inorganic Hg input, transport, and methylation potential (Wyn et al. 2009; Gabriel et al. 2014; Gorski et al. 2008; ; Chaves-Ulloa et al. 2016; Chételat et al. 2018; Rudd et al. 2018; Schartup et al. 2018). The complex chemical conversions and cycling of Hg make it particularly challenging to predict the concentration of MeHg in fish and wildlife from concentrations of inorganic mercury in air, water, and sediments (Gustin et al. 2016, Sunderland et al. 2016; Eagles-Smith et al. 2018). Importantly, even in areas where Hg deposition is low, concentration in biota may be disproportionately high if conditions are conducive to MeHg production and biomagnification.

The selection of monitoring locations (which could include the multi-level approach of primary and secondary sites) will need to account for the broad geographic range of methylation abilities in oceanic and continental areas, while also responding to the primary monitoring questions (see 6.3.5. for more details). The response from one site is not necessarily relatable to the response of a neighboring site that has different habitat characteristics. And, how a collection of habitats within an ecosystem respond to mercury input will play a role in selection of monitoring locations according to the information questions of interest (see section 4.4 on evaluating the sensitivity of ecosystems to Hg input). Once monitoring locations are chosen, tracking temporal trends will be possible by performing consistent sampling over multiple years.
4.4. Ecosystem sensitivity and its importance to assessing threat

Ecosystems are extremely variable in their relative sensitivity to mercury methylation. This is largely due to the heterogeneity of abiotic and biotic processes that influence the ability of any particular ecosystem to convert available inorganic mercury into its more bioavailable organic form (via the methylation process). Understanding this variability (with a particular interest for highly sensitive areas) is important during the process of identifying monitoring locations especially when addressing the operational questions (from Table 2.1) to characterize spatial patterns and estimate exposure and adverse impacts. Conversely, selecting monitoring locations that are less sensitive may be important for tracking temporal trends to reduce confounding variables. Therefore, a mix of monitoring locations that represent both sensitive and less-sensitive ecosystems can address multiple questions is viewed as most useful and identifying the key variables that may provide some direction for selection are important.

The amount of total mercury in any given location does not necessarily correlate to ecosystem impact – largely due to the complexities of the methylation process. Ecosystems that are highly sensitive to mercury methylation may require only limited amounts of mercury to pose risks to organisms. Similarly, ecosystems with little to no sensitivity to mercury methylation may experience high levels of mercury inputs with limited impacts to the environment and human health. To credibly assess the potential threat of mercury to biota, biodiversity, ecosystem services and people, it is important to combine the two factors of mercury risk (from multiple inputs) and ecosystem sensitivity (to converting available mercury into its more toxic, bioavailable form). The structure of this approach to mercury threat assessment and the associated information is provided in Table 4.3 and can be potentially used as a criterion tool for selecting monitoring locations.

Table 4.3. Description of mercury threat assessment categories, key variables and comments on indicators being used for each.*

<table>
<thead>
<tr>
<th>Threat assessment category</th>
<th>Key variables</th>
<th>Indicator comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystem Sensitivity</td>
<td>Land cover types</td>
<td>There are multiple land cover types – each will be weighted according to their relative methylation potential. For example, low-moisture deserts have little methylation potential in comparison to tropical forests.</td>
</tr>
<tr>
<td></td>
<td>Wetland types</td>
<td>There are multiple wetland types – they are often nested within land cover types. Each will be weighted according to their relative methylation potential. Lakes, ponds, and rivers will be assessed (knowing that their shorelines with wetlands have great methylating abilities), as will bogs, peatlands, and swamps; thawing permafrost can create anoxic methylation hotspots compared with frozen peat plateaus.</td>
</tr>
<tr>
<td></td>
<td>Habitat characteristics</td>
<td>Additional factors related to habitats that influence methylation potential that are not mapped by land cover or wetland type. For example, areas with higher soil organic carbon (SOC) can have higher methylation potential than areas with lower SOC levels.</td>
</tr>
<tr>
<td>Mercury Inputs/Risk</td>
<td>Volcanic activity</td>
<td>Ecosystems sensitive to Hg input may be impacted by natural sources from local volcanic or thermal vent activities.</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ASGM activity</td>
<td>Artisanal and Small-scale Gold Mining (ASGM) – is the largest source of global mercury emissions. ASGM activity is highly variable in its distribution and relative amount of activity, and is of global and local importance.</td>
<td></td>
</tr>
<tr>
<td>Power Generation</td>
<td>Stationary Power Sources – Globally, stationary power sources contribute the most mercury emissions of all large-scale industrial activities. Of these stationary power sources, coal-fired power plants are the largest single contributor. Quantifying the amount of industrial activity within watersheds is an important indicator of local mercury emission potential.</td>
<td></td>
</tr>
<tr>
<td>Water-level Management</td>
<td>Reservoirs – The creation of reservoirs creates a pulse of mercury through the release from soils and drowned vegetation and can last 1-2 decades. The management of water levels thereafter can further exaggerate the shoreline methylation process through frequent wet-dry cycles that can lead to increases in methylation exposure to aquatic and terrestrial biota.</td>
<td></td>
</tr>
<tr>
<td>Large-Scale Industry</td>
<td>Large Scale Mining (LSM) – LSM is the largest mercury emitter of all large-scale industrial activities, with non-ferrous metal production at the top of the list. Cement production is a close second. Tracking the amount of large-scale industrial activity in each watershed is an important indicator of mercury contamination of air and water.</td>
<td></td>
</tr>
<tr>
<td>Intentional Uses</td>
<td>Waste Disposal – Emissions associated with mercury-added product waste is the largest source of emissions in this sector. Documentation the location of waste disposal sites within watersheds is important for guiding monitoring of potential contamination.</td>
<td></td>
</tr>
<tr>
<td>Agricultural activity</td>
<td>Rice paddy fields – certain agricultural activities, such as planting rice in paddy fields, can change Hg input. Rice paddy fields are a dominant agricultural land use throughout Asia and have been identified as important sites for methylmercury production in the terrestrial ecosystem and a primary pathway of MeHg exposure to humans in mercury mining areas.</td>
<td></td>
</tr>
<tr>
<td>Degradation caused by unsustainable land management practices</td>
<td>Deforestation – is likely the most important process driving mercury releases to the water in tropical forest regions. Incorporating the amount of deforestation in a watershed is an important indicator of disturbance and potential mercury releases.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Soil Erosion – is the primary process that carries mercury from the land into freshwater ecosystems. There are many factors influencing soil erosion, and it is responsible for releasing mercury into the air and water. It is particularly pronounced</td>
<td></td>
</tr>
</tbody>
</table>
where ASGM activity and deforestation occur but is not limited to these areas. Soil erosion is a good proxy for habitat degradation, and an important indicator of the mercury transport process in terrestrial and freshwater environments.

**Fire Frequency** – Fires are a natural disturbance process in many ecosystems. However, the frequency and intensity of fires has been influenced by climate change in many ecosystems, including forests and wetlands in the tropics. Fires result in the natural release of mercury into the air, and the more fires there are, the more mercury is likely emitted.

* Mercury inputs are categorized according to UNEP (2019) nomenclature. Degradation caused by unsustainable land management practices have been included as they also have a direct influence on the availability of mercury in the environment.

The information available to create a robust global threat assessment requires combining a mix of discrete categorical and continuous data. Methods are being developed that ensure both consistency and transparency in this approach, as well as the ability to down-scale this approach for application at regional and local levels to make use of critical information not available at a global scale (e.g., point-source data). As water is a major pathway for mercury through ecosystems, and watersheds offer a justifiable, hierarchical approach to assessment across many spatial scales, evaluating the threat of mercury via watersheds has emerged as an important part of this approach (Evers and Sunderland 2019). Creating new assessments of watershed risk, sensitivity and threat of mercury impacts to nature and people will significantly improve the selection of priority sites for global mercury biomonitoring that will most effectively use limited resources. Information from these biomonitoring priority areas can, in turn, be used to adaptively manage and improve the usefulness of mercury threat-related assessments over time. This supports the application of “systems-thinking” considered necessary to chemicals and waste problem-solving in which “a set of synergistic analytical skills is used to improve the capability of identifying and understanding systems, predicting their behaviours, and devising modifications to them to produce desired effects” (Arnold and Wade, 2015).

While the precision of modelling the sensitivity of ecosystems is still in its early stages and therefore introduces its own uncertainty, there will be a level of useful contributions for decision-making such as prioritization of sites for the Effectiveness Evaluation of the Convention. At its most basic level, it will be possible to at least categorically identify ecosystems or areas that tend to have relatively lowest (e.g., deserts) or highest sensitivity (e.g., wetlands), and low or high inputs (e.g., from global Hg deposition).

### 4.5. Ancillary measurements

Ancillary measurements often collected with biota mercury data include species, total body length, weight, and spatial coordinates (low resolution level) and in addition fat levels, and stable isotopes of carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) and other variables can be included (high resolution level). Stable isotope measurements in biota assist with identifying changes in food web structure and trophic position and feeding habitat (Abeysinghe et al. 2017) and aid in evaluating causes of temporal trends in the context of abiotic factors such as changing air emissions, sediment and water chemistry, and temperature. Without these ancillary measurements, and analyses that normalize data in the context of food web dynamics, it will be challenging to determine if the observed changes, or lack thereof, is due to
changes related to the efforts related to the Minamata Convention or driven by large-scale factors such as changes in food web complexity, trophic position of biota, climate change, overfishing, and biogeochemical conditions. AMAP’s mercury monitoring programs for biota include these ancillary measurements and surveillance efforts conducted by Canada and Norway also sometimes include stable isotope measurements of carbon ($\delta^{13}$C) and nitrogen ($\delta^{15}$N) (Braune et al. 2015, 2016). A range of low- and high-resolution approaches to collect ancillary data can be maintained depending on the needs of the monitoring program (Table 4.4).

### Table 4.4. Example of low- and high-resolution levels for sampling and analysing biota in conjunction with ancillary measurements.

<table>
<thead>
<tr>
<th>Resolution Level</th>
<th>Matrix</th>
<th>Ancillary measurement examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Resolution</td>
<td>Biota (e.g., muscle tissue)</td>
<td>Species, body length, mass, spatial coordinates</td>
</tr>
<tr>
<td>High Resolution</td>
<td>Seawater</td>
<td>pH, dissolved oxygen, salinity, temperature, depth, mercury ($\delta^{202}$Hg) stable isotopes</td>
</tr>
<tr>
<td></td>
<td>Surface Sediments (e.g., top 2 cm)</td>
<td>Carbon ($\delta^{13}$C), nitrogen ($\delta^{15}$N), mercury ($\delta^{202}$Hg) stable isotopes, depth, temperature</td>
</tr>
<tr>
<td></td>
<td>Biota (e.g., muscle tissue, blood, keratin-based tissue)</td>
<td>Species, body length, mass, spatial coordinates, carbon ($\delta^{13}$C), nitrogen ($\delta^{15}$N), mercury ($\delta^{202}$Hg) stable isotopes</td>
</tr>
</tbody>
</table>

It is likely that a tiered approach (see 6.3.5) will be beneficial for evaluating the effectiveness of the Minamata Convention when considering the existing and new data for biota as they will have different levels of resolution and quality assurance and quality control (QA/QC).

Specifically, body length, mass, species name, and spatial coordinates (latitude/longitude) are nearly always collected as metadata in mercury monitoring programs for biota. However, some studies also collect data from other matrices including seawater and marine sediments (Azad et al. 2019b) along with high resolution ancillary variables including but not limited to carbon ($\delta^{13}$C), nitrogen ($\delta^{15}$N), and mercury ($\delta^{202}$Hg and $\delta^{199}$Hg) stable isotopes (Cransveld et al. 2017), pH, salinity, sea depth, organic carbon, dissolved oxygen, temperature, etc.

Braaten et al. (2019) argue that to evaluate the effectiveness of the Minamata Convention, there is a need for quantification of legacy Hg sources, taking into account their bioavailability, and for separating these sources from long-range atmospheric sources of Hg. Measurement levels and fluxes from abiotic media, such as water and sediment/soils, could be included in Tier 3 biomonitoring to help quantify legacy sources and provide further support in understanding temporal trends and spatial patterns. Abiotic media should not be used exclusively because of interpretative limitations and uncertain connectivity with associated biota, especially high trophic level species.

For each terrestrial location, this should include lake and catchment morphology, pollution deposition patterns, and local pollution history. For each animal species data must include length, weight, sex, and age (when it can be obtained). Samples (i.e., fish muscle) for determination of total Hg concentrations, could also be analysed for stable isotopes of nitrogen ($\delta^{15}$N) and carbon ($\delta^{13}$C) for a better understanding of trophic position and energy sources.
4.6. Tiered approach to monitoring

For the Effectiveness Evaluation, an integrated tiered approach to monitoring of mercury in biota is being proposed (see chapter 2).

For mercury monitoring in biota, a low resolution-level set of parameters can be defined (e.g., precise spatial coordinates of sampling site), fish Hg concentration (muscle, wet weight; fish sampling restricted to the same season), fish weight and length; preferably fish age, sex and maturity stage, and supplementary information on the lake and catchment (i.e. size, elevation, and land cover and use (natural, agriculture, developed)). Higher resolution interests would include time series that would preferably be established for lakes with well-known pollution loading (local catchment sources, or from long-range transported air pollution). Inclusion of lakes with only external (long-range transported) pollution loads is crucial for effect evaluation in remote areas (boreal, subarctic, arctic). In addition to low resolution parameters further ancillary data should be included such as stable carbon and nitrogen isotopes for fish, water chemistry (including Hg species, total and methyl-Hg, TOC, pH (all mandatory), and nutrients (preferred)). For high resolution scenarios, information on diet (from fatty acids, for example), stable isotopes of lower foodweb organisms, to provide data on food web structure.

The tiered approach would build upon the use of existing monitoring networks and stations. To best represent global patterns related to both local and long-range transport of mercury, additional monitoring stations could be needed to represent ecosystem sensitivity/insensitivity (which can be mapped with a certain level of value for decision-making) and have a mixture of background/reference sites together with sites with well-known local Hg sources. Bioindicators can be identified that are cost-effective and replicable over time. A proposed tiered approach for biota monitoring, which would be supported by existing and new programmes, is shown in the Annex to this guidance.

4.7. Implementation of the tiered approach

In view of the current gaps in mercury monitoring data in biota, the need to understand temporal trends, spatial patterns, source types and their ecosystem linkages, and the interest of using bioindicators relevant for human exposure assessment and ecological health, two overarching biotic Hg monitoring components – one for landscapes (continents) and another for seascapes (oceans) – provide potential fundamental structure to implement the tiered approach and support global biomonitoring efforts under the Minamata Convention.

The oceanic and continental monitoring components would integrate existing and ongoing data collection efforts with new sampling with a view to providing relevant information in a cost-effective way. This would be achieved by building upon existing long-term efforts and identifying data gaps so further global assessments may be conducted. Below is a brief overview of the main aspects of two proposed frameworks. Further information may be found in the technical background document on biota monitoring (Evers and Sunderland 2019). This potential structure further contributes to the need for developing comparable data for Parties and relevant organizations for biomonitoring within their own programmes.

4.7.1. Oceanic component

The cycling and movement of Hg in the world’s oceans varies by hemisphere, basin and juxtaposition with the continental land masses (Sunderland and Mason 2007; Mason et al. 2012; Zhang et al. 2020).
Therefore, Hg concentrations in fish, sea turtles, birds, and marine mammals vary significantly across geographic areas.

The proposed oceanic component of global monitoring of mercury in biota is composed of three steps, each further divided in three sub-steps. Step 1 starts with a characterization of ocean basins and is followed by collecting data on fish and other biota from existing mercury monitoring programmes. Step 2 is focused on the selection of key bioindicators (e.g., Drevnick et al. 2015, 2017), and Step 3 focuses on analysing existing data, determining optimal sample sizes, and collecting and analysing samples (Figure 4.2; Evers and Sunderland 2019).

**Figure 4.2.** Potential stepwise components for biomonitoring in oceans.

### 4.7.2. Continental component

To understand the complexities of a landscape and its ability to methylate mercury and make it available in the food web, biota monitoring under the Continental Framework would require multiple defined steps (Figure 4.3). In this proposed framework, Step 1 would focus on mapping the best locations for global Hg monitoring, with a focus on wetlands and other sensitive ecosystems. Mercury methylation is highest in wetlands – and, potentially greatest in estuarine wetlands such as mangroves, and peatland bogs that are generally acidic. Forested areas are also an important habitat for increasing dry deposition rates of atmospheric Hg in temperate (Driscoll et al. 2007) and tropical (Gerson et al. In Review) ecosystems, while agricultural areas tend to dampen methylation rates (Chen et al. 2008) with the exception of rice paddies (Abeyesinghe et al. 2017). Step 2 would focus on identifying overlap in existing data with areas that are of particular relevance for monitoring, including ASGM sites, important freshwater fishing areas and key biodiversity areas, and identify potential bioindicators for harmonized sampling and comparison (example of global fish study compared multiple fish species from a broad variety of locations; Buck et al. 2019). Step 3 would select the best ecosystem sensitivity spots based on the two previous steps to establish spatial patterns and temporal trends, as well as linkages to Hg sources.

---

**Step 1**
- a. Identify physical distinctions among ocean basins of interest;
- b. Collect FAO commercial fisheries data or national data for targeted ocean basins;
- c. Designate useful existing programs and mercury data.

**Step 2**
- a. Identify tuna, billfish, and other fish populations of greatest commercial and recreational interest by ocean basin - for open ocean biomonitoring;
- b. Identify barracuda and mahi mahi populations for coastal ocean biomonitoring;
- c. Identify other trophic level 4 bioindicators, including marine mammals, sea turtles and birds, that are relevant to the Minamata Convention.

**Step 3**
- a. Determine existing mercury data overlaps of best bioindicators for ocean basins of interest;
- b. Conduct a power analyses based on the species/groups selected and their known mercury concentrations within that ocean basin to determine sample size;
- c. Collect samples from target bioindicators with an emphasis in areas with data gaps.
Field sampling and laboratory analysis depend on a number of factors, including the taxa chosen, their habitats and abiotic conditions, objectives of the monitoring, etc. Protocols for collecting samples from biota are available for all tissue types, and examples of the proper selection of tissue type are well-established with associated information about the percent methylmercury content in the tissue and the preferred type of tissue preparation. Most muscle, blood, egg and keratin-based (e.g., scutes, feathers, and fur) tissues primarily contain methylmercury. This is important for the simpler and more cost-effective laboratory analyses of total mercury concentrations that can assume 95% or more methylmercury content (with some exceptions).

The timing of biota sampling at monitoring locations varies according to the objectives, habitats/ecosystems, and chosen bioindicators. The fraction of mercury retention in the atmosphere, soils, and waters can vary over days to centuries (Figure 4.4). Therefore, knowledge of mercury retention in habitats that biota is sampled from will be important for understanding spatial patterns, temporal trends, and linkages with mercury sources. Sample timing also depends on the rate of change in Hg concentrations in the bioindicator tissues of choice.

Information on climate variables, habitat type, and taxa ecology are generally needed for proper interpretation of spatial and temporal patterns. Likewise, knowledge of mercury retention in habitats that biota is sampled from will be important for understanding spatial patterns, temporal trends, and linkages with mercury sources. For linkages to mercury source types, mercury isotopes are important.

To understand mercury exposure and the potential effects on taxa, it is important to know the age category, morphometrics (e.g., weight, length, etc.), and foraging ecology.

In some cases, where total mercury body burden changes rapidly, such as in fish and birds within lakes with small watersheds (Evers et al. 2007), changes can be detected on the scale of years (Wiener et al.)
Biomonitoring in areas with smaller changes in environmental loadings, but with more complex ecosystems that contain varying processes that sequester, and methylate mercury require sampling every other year for one or two decades (Riget et al. 2011; Eagles-Smith et al. 2016a; Sunderland et al. 2018; Evers et al. 2020). Within ocean basins, increasing mercury concentrations were detected over multiple decades for tuna in the Pacific Ocean (Drevnik et al. 2015; Drevnik and Brooks 2017).

Figure 4.4. Retention of mercury fraction (0 to 100%) over time (days to centuries) in various compartments of the atmosphere, landscape (e.g., soils), and waterscape (e.g., ocean waters and sediments). Source: Amos et al. (2014).

4.8.1. Quality assurance and quality control for biota monitoring

The quality control and quality assurance of mercury concentrations analyzed from various types of animal tissue are important and require proper standard reference materials and the use of duplicate sample analyses and blanks.

While instrument calibration is important for obtaining accurate and comparable mercury data from biota, especially when comparing different types of instruments (e.g., DMA vs. CVAA), sample handling and processing are by far the greatest sources of introduced variability in mercury levels in animal tissues. Therefore, protocols for sample preparation will need to be carefully vetted, described, and followed.

Furthermore, similar to other matrices, analytical operations for biota monitoring will also need to follow strict chain-of-custody and standard operating procedures for laboratory analysis and data handling.

Further information on quality assurance and quality control is available as Supplementary Material to this guidance.
**4.8.2. Statistical considerations**

A wide array of statistical tests are available to evaluate temporal trends of mercury levels in biota including, but not limited to, general linear (GLM) and non-linear (e.g., logistic regression) models, classification and regression tree (CART), Mann-Kendall (MK) test, and Bayesian model selection and uncertainty assessment techniques including the widely used Akaike information criterion (AIC), etc. For evaluating spatial trends GLM, general additive modelling (GAM), kriging or Gaussian process regression, Cox point process and spatial covariance modelling, principal component analyses (PCA), multiple-response permutation procedures (MRPP), probability density estimations and Monte Carlo simulations are some of the approaches that can be used for existing or new biota data sets. While statistical tests may inform optimal sample sizes, power analyses combined with probability interests and variability of mercury concentration in different tissues are a more suitable basis for choosing the type of sample to be collected (see above).

Length or body mass normalization of biota will be critical for interpreting mercury and methylmercury data. Moreover, evaluating trophic position and food web structure using carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$) in conjunction with mercury ($\delta^{202}Hg$) stable isotopes and other matrices such as seawater and sediment can support more rigorous high-resolution modelling although targeted sampling may be required to achieve this goal.

As with any analysis, when dealing with biota monitoring in the context of the Effectiveness Evaluation, there are important limitations and uncertainties that need to be conveyed in a clear and transparent way. Specifically, there are major challenges linking mercury levels in biota with mercury concentration in abiotic matrices such as air and water especially considering post depositional processes, trophic position, changes in food web structure and complexity, and broad-scale drivers such as environmental chemistry factors (e.g., pH, DOC), temperature, geography, species growth rates, and climate change (Braune et al. 2015, 2016).

**4.9. Conclusions**

Biota monitoring data can help address the operational questions (Table 2.1) and support the Effectiveness Evaluation of the Convention.

Historic data available from various biota monitoring programmes, databases and other resources can be used to improve our understanding of the exposures to mercury in biota before the Minamata Convention’s entry into force and to help establish a baseline for the Effectiveness Evaluation. Moving forward, existing government-led national mercury monitoring programs, regional initiatives, and/or academic-led studies can provide comparable biota monitoring data for use in the Effectiveness Evaluation. New monitoring efforts may further contribute by providing comparable data on key bioindicators filling data gaps, and building capacity.

Biota monitoring data can be collected with focus on continental and oceanic components designed as part of a Tiered approach for Parties and organizations who elect to develop new monitoring programs or improve existing ones. Briefly, Tier 1 is on those seeking to create a biota-based monitoring program, or expand a minimal program, but that may not have sufficient resources to implement the actions in Tier 2. The goal of Tier 1 should be to focus on temporal trends and collect total mercury measurements from trophic level 3 or 4 biota that best represent the targeted habitats Tier 1 activities should ideally be repeated for the same species using the same size classes in the same habitat every 2-5 years. Tier 2 aims to realize information that will contribute more meaningfully to the Effectiveness Evaluation, and
thus calls for more in-depth analysis of the Tier 1 monitoring efforts, or incorporation of mercury biomonitoring into other, in-depth mercury monitoring efforts. Tier 3 aims to increase understanding of key processes that link mercury sources to key bioindicators, and thus resource-intensive research methods and approaches are required.

Key elements that are essential to all monitoring efforts for biota include: a) defining the target bioindicators and sample size, which usually focus on high trophic level biota that are vulnerable to relatively high methylmercury exposure; b) selecting and measuring the appropriate biomarkers (i.e. tissue types) to best interpret exposure to different sources and forms of mercury, with total mercury measurements in muscle tissue of fish and marine mammals, blood, and feathers or eggs of birds being most commonly used and accepted; c) identifying the monitoring locations that best reflect the objective for biomonitoring (e.g., temporal, spatial, or ecological health questions) through the use of an ecosystem sensitivity modelling tool; and d) managing and analysing data as per the guiding operational questions for the Effectiveness Evaluation. All these aspects can use well-established standard operating procedures available in the Supplementary Material.
Chapter 5. Human Biomonitoring

5.1. Introduction

Understanding human exposures to chemical hazards through biomonitoring activities is important for scientific and regulatory purposes (WHO Regional Office for Europe 2015; Louro et al. 2019). For mercury, in particular, human biomonitoring practices (i.e., mercury measures in hair, urine, and/or blood) are well-understood, practiced by some national governments, and can help assess the efficacy of policy actions (WHO 2018a; UNEP 2019; HBM4EU 2019).

The recent Global Mercury Assessment 2018 showcased biomonitoring efforts worldwide, and in doing so illustrated the diversity of efforts ranging from engagement of vulnerable communities situated in remote and resource-limited settings to national-level surveys implemented by government agencies involving thousands of participants (UNEP 2019). Human biomonitoring of mercury is relatively uncomplicated; these measurements are scientifically sound, technically simple with validated protocols available, and can be conducted at relatively low cost (Evers et al. 2016).

Human biomonitoring data can help address operational questions that support the Effectiveness Evaluation (Table 2.1). First, quality measures of mercury levels in human biological samples (herein referred to as biomarkers) provide direct evidence of exposure in a given population at a given time. Second, such measures, when coupled with questionnaire data, may offer insights into possible sources and routes of exposure from which attributions may be deduced. Third, temporal changes can be gleaned if monitoring is repeated in the same population over time. Fourth, biomonitoring data can be inputted into established risk assessment frameworks to estimate health impacts including burden of disease, as well as to assess the efficacy of different risk management strategies. The operational questions that support the Effectiveness Evaluation can provide the foundation to design a human biomonitoring study (that uses existing data and/or purposefully produces new biomonitoring data), and guidance for realizing this is detailed below.

Successful human biomonitoring activities require a multi-disciplinary and inter-sectoral team to work collaboratively across all aspects of the effort, from setting research questions that guide the design of biomonitoring activities to the interpretation and communication of results (Figure 5.1). Information in this chapter provides essential guidance (and links to key resources) for Parties and relevant organizations to consider in terms of using existing, and generating new, human biomonitoring data for the Effectiveness Evaluation. This chapter also provides a brief overview of our state of knowledge for human biomonitoring of mercury, proposes a framework by which biomonitoring data can be used for evaluating the effectiveness of the Convention, and then offers guidance on best scientific practices to:

a) define the target and sample population; b) select and measure the appropriate biomarkers to help tease apart exposure to different sources and forms of mercury; c) administer surveys to gather supportive information to deepen understanding; and d) manage and analyze data as per the guiding policy question. All these aspects must be performed in a responsible and ethical manner. While the focus here is on Article 22 (Effectiveness Evaluation), many of the details below synergize with other articles of the Convention (e.g., Articles 4, 7, 14, 16-19, 21).
58

5.2. State of knowledge

5.2.1. Existing data

To assess our current understanding of human exposures to mercury, a systematic search of the recent (2000 to 2018) literature identified 312 studies from 75 countries from which 424,858 mercury biomarker measurements from 335,991 individuals were analyzed (Basu et al. 2018). This activity was sponsored by the World Health Organization (WHO) as part of the Global Mercury Assessment 2018 (UNEP 2019). The authors of this report concluded that blood, hair, and urine mercury levels are generally less than 5 μg/L, 2 μg/g, and 3 μg/L, respectively, in background populations with no significant sources of exposure to mercury. The results also identified populations with elevated exposures. From this dataset there are two key groups of human biomonitoring data to be aware of.

First, national human biomonitoring programs exist that aim to derive information that is representative of a country or region. These are usually sponsored and/or operated by government agencies, are resource intensive, and generally cover many chemicals. These studies therefore tend to use random sampling of an adequate population size and use reference laboratories for mercury analysis. Sample sizes range from a few hundred to several thousand. The Global Mercury Assessment 2018 human
The biomonitoring dataset contains 192,651 biomarker measures from these programs. However, national biomonitoring programs that consider mercury exposure are only carried out in 9 countries to date, and international representation is mostly limited to higher income regions.

Second, there exist data (i.e., 232,207 biomarker measures) from cross-sectional and birth cohort studies. The design and quality of these studies vary tremendously. Further, the sample populations usually are not representative of the target population as most rely on convenience sampling. Nonetheless, these studies are of importance as they tend to focus on vulnerable groups identified by the Minamata Convention (e.g., women of child-bearing age). Also, some of these efforts exemplify how mercury human biomonitoring may be performed successfully on a regional basis, such as the Arctic Monitoring and Assessment Programme (AMAP) and the DEMOnstration of a study to Coordinate and Perform Human biomonitoring on a European Scale (DEMOCOPHES) effort.

5.2.2. Existing data gaps
Despite current understanding of human exposures to mercury worldwide, there is great variability in exposures around the world and across/within population groups. Arguably the greatest data gap concerns the many countries and regions without any mercury biomonitoring data without which evidence-based decision making is hampered. Notably, nearly 70% of the data in the Global Mercury Assessment 2018 biomonitoring dataset was represented by just 8 countries (Republic of Korea, China, Japan, United States, Brazil, Saudi Arabia, Canada, and the Russian Federation).

5.2.3. Future data sources
We can expect, with very high confidence, that mercury human biomonitoring data will be available in the future from two primary areas. First, some national human biomonitoring programs are firmly established by governments with sampling frequencies every 1-2 years (e.g., Canadian Health Measures Survey (CHMS), Czech Republic Environmental Health Monitoring System (EHMS), German Environmental Survey (GerES), Republic of Korea’s National Environmental Health Survey (KoNEHS), US National Health and Nutrition Examination Survey (NHANES)), and these will be dependable programmes for evaluating the effectiveness of the Convention. Second, future data may also be expected from cross-sectional and birth cohort studies. Though, these are largely ad-hoc efforts run by academic researchers who depend on extramural funding, and as a collective they are not purposefully designed nor coordinated to address long-term effectiveness evaluation. It is also noted that many existing human biomonitoring programs, not necessarily designed for mercury exposure assessments, collect and archive blood samples (and other matrices) that may be analyzed retrospectively.

A third way forward, and in particular to help fill data gaps in a globally coordinated manner, Parties and relevant organizations without existing data sources should consider, where possible, a harmonized approach to launch new biomonitoring studies. A good starting point is the recent guidance from the WHO to characterize prenatal mercury exposure (World Health Organization (WHO) 2018a). Using this WHO protocol would enable the collection of comparable data (e.g., samples from 250 individuals per a defined study location, with minimum diversity recommended), through addressing the most vulnerable population group i.e. the fetus. The studies would be country driven such that local ethical clearance would be required, and the studies would be conducted within the national health system. With funding from the Global Environment Facility (GEF), under the project “Develop a Plan for Global Monitoring of
Human Exposure to and Environmental Concentrations of Mercury”, this WHO protocol was piloted between 2015 and 2017 in diverse settings and several countries. Examples of diverse human mercury exposure sources targeted in this WHO project included rice consumers (in China), seafood consumers (in Ghana and India), local industrial contamination (in India), mercury primary mining (in Kyrgyzstan), artisanal and small-scale gold mining (ASGM, in Mongolia), and freshwater fish consumers (in Russia). The GEF project showed that the generation of data using the WHO protocol in low- and middle-income countries is cost-effective, practical, and feasible. The project also built local capacity to conduct relevant studies, which can therefore be repeated over time and in a range of locations to fill gaps.

5.3. Proposed framework

This section outlines a proposed framework in which monitoring programmes can provide comparable human biomonitoring data for the Effectiveness Evaluation. Driven by questions that support the Effectiveness Evaluation (Table 2.1), there are two main components to the proposed framework to bear in mind:

Pre-Minamata Convention period: 1) the use of existing biomonitoring data contained in the WHO-sponsored, Global Mercury Assessment 2018 biomonitoring dataset, or from other existing sources, can be used to help understand human exposures to mercury before the Minamata Convention’s entry into force (i.e. help establish the baseline).

Effectiveness Evaluation period: 2) the use of biomonitoring data expected in the future from government-led national biomonitoring programs, regional initiatives, and/or academic-led studies; and 3) implementation of new biomonitoring studies led by Parties and relevant organizations in a harmonized way so that they are purposefully designed to fill data gaps, build capacity, and support the Effectiveness Evaluation. During the first Effectiveness Evaluation period, human biomonitoring activities may be designed according to the tiered approach outlined below.

The biomonitoring data collected from such activities: a) provide direct evidence of mercury exposure in a given population at a given time; b) when coupled with questionnaire data, offer insights into possible sources and routes of mercury exposure from which attributions may be deduced; c) can assess temporal changes in mercury exposure if monitoring is repeated in the same population over time; and d) assess potential health impacts and contribute to risk management activities.

The guidance presented below is intended to be fit for purpose i.e., Minamata Convention stakeholders with narrow (e.g., specific country, population, or hotspot) or broad (e.g., global understandings, long-term trends) interests can generate comparable data to address the same relevant questions, albeit on different scales.

5.4. Tiered approach for human biomonitoring

Mercury human biomonitoring data can be designed as part of a Tiered approach for Parties and relevant organizations who may wish to improve existing monitoring programmes, or develop new programmes, with a view to contributing to the Effectiveness Evaluation. Details of the Tiered approach are summarized in the annex below.

---

12 UNEP/MC/COP.3/INF/19.
Tier 1 – For Parties and organizations seeking to create a human biomonitoring program, or expand a minimal program, but that may not have sufficient resources to implement the actions in Tier 2, the goal should be to focus on a vulnerable sub-population (section 5.6) and take total mercury measurements in blood, urine, or hair (section 5.7). This activity should ideally be repeated in the same population every 2-5 years. A good starting point for Tier 1 is the recent guidance from the WHO to characterize prenatal mercury exposure (WHO 2018a).

Tier 2 – Building on Tier 1 activities, Tier 2 biomonitoring activities will perform more in-depth analysis of the Tier 1 sub-population group (e.g., measure total mercury in blood, urine and/or hair; consider measuring methylmercury and/or mercury stable isotopes in these biomarkers), or incorporate mercury biomonitoring into other, in-depth health surveys or cohort studies. These activities are more expensive and complex than those under Tier 1, but they provide information that will address all operational questions to support the Effectiveness Evaluation (table 2.1).

Tier 3 – To increase understanding of key processes that link mercury sources to human exposures, resource-intensive research methods and approaches are required. These include national human biomonitoring programs, or careful design of Tier 2 activities with coordinated air and biota sampling.

5.5. Ethics

It is imperative that human biomonitoring activities adhere to the World Medical Association’s Helsinki Declaration, and that proper ethical approvals are in hand before any human subject research occurs. In most countries, Ministries of Health along with tertiary academic institutions, are the primary contact point for obtaining such ethical approvals. In some countries, sub-national/regional governments have self-determination of research activities and their own ethical guidelines and research licenses need to be followed, for example the National Inuit Strategy on Research. Moreover, depending on the national context, specific organizations (e.g., workers unions, occupational safety boards, industry groups, dental/medical associations) may also have ethical guidelines to follow.

Given that human biomonitoring may focus on vulnerable populations, participatory engagement of pertinent stakeholders (e.g., study participants, workers, community leaders, health care providers, regional authorities) is necessary not only for ethical and safety purposes but to also help ensure that the best studies are designed, conducted, and communicated. The “International Ethical Guidelines for Health-related Research Involving Humans”, prepared by the Council for International Organizations of Medical Sciences (CIOMS) in collaboration with WHO, should be consulted (CIOMS, 2016). In addition, Parties and organizations may consult literature on legal, ethical, and social issues pertaining to human biomonitoring from the European Human Biomonitoring Initiative (HBM4EU, 2018a), the Canadian Health Measures Survey (Day et al., 2007), the International Labour Organization’s Technical and Ethical Guidelines for Workers’ Health Surveillance (ILO, 1998), and the World Health Organization’s recent guidance on ASGM (WHO, 2021a).

With regards to data ownership, human biomonitoring activities must respect the legislation of individual countries and this may vary depending on the population that is being sampled. For example, in Canada, Indigenous communities own the human biomonitoring data collected in their community (i.e., OCAP Principles – Ownership, Control, Access, and Possession), instead of the data being owned by

13 Available at https://www.itk.ca/national-strategy-on-research-launched.
the country or the organization responsible for generating the data. Appropriate communication and
dissemination of data results back to the contributors is another important aspect of human
biomonitoring. Moreover, in ethical research, all participants have the right to withdraw from
studies/monitoring and have all their data and samples removed from the data set and no longer used.

5.6. Human population group

5.6.1. Identification of target population

All human populations worldwide are exposed to some amount of mercury (UNEP and World Health
Organization (WHO) 2008; Basu et al. 2018). There is thus value in assessing mercury exposures in both
the general population as well as in vulnerable groups. The selection of a specific target population will
be guided by the interests of the Parties or relevant organizations carrying out the monitoring activities,
in consideration of operational questions that support the Effectiveness Evaluation (chapter 2). For
example, some initiatives may choose to focus on the general population while others may choose to
focus on a specific vulnerable group (e.g., pregnant women, workers and community members living
around ASGM sites, Indigenous Peoples, and local communities).

In terms of evaluating mercury exposures in the general population, the geographic scope (e.g., discrete
community, entire country) and sociodemographic profile (e.g., sex, age) of this target population needs
to be defined a priori. For guidance on studying general populations, Parties and relevant organizations
can refer to aforementioned national human biomonitoring programs that tend to have detailed
protocols available.

In terms of evaluating mercury exposures in population groups most vulnerable to mercury exposure,
there are two broad groups to consider. First, early lifestages (i.e., fetus, newborn and children) are
susceptible to mercury exposure because of the sensitivity of the developing nervous (and other
physiological) system. This population group can also include pregnant women and/or women of child-
bearing age. Second, some populations are vulnerable because they are exposed to higher levels of
mercury. A resource document to help identify sub-populations that may be at risk of mercury exposure
and health impacts was produced through a collaboration between UNEP and WHO (WHO 2008).

Human exposure to elemental and inorganic mercury may occur in occupational settings (e.g., ASGM
and dentistry practices), from contact with certain products (e.g., dental amalgams, some skin-lightening
creams, broken fluorescent bulbs and other waste products), and from environmental contamination

Human exposures to organic mercury largely arise from dietary sources. Mercury released into the
environment may be converted by microorganisms to methylmercury which bioaccumulates and
biomagnifies through the food web, particularly in aquatic systems (see chapter 4). Sampling of
freshwater fish and seafood has found widespread methylmercury contamination, with some widely-
consumed predatory species, such as tuna, swordfish, grouper, and mackerel being among the most
highly contaminated.\textsuperscript{14} Therefore, for many population groups, dietary consumption of contaminated
fish, shellfish, and marine mammals is an important source of exposure. Seafood, however, is the main

\textsuperscript{14} Global Environment Monitoring System (GEMS) / Food Contamination Monitoring and Assessment Programme,
available at: https://www.who.int/teams/nutrition-food-safety/databases/global-environment-monitoring-
system-food-contamination
source of protein and nutrients for billions of people worldwide (FAO 2020). Other staple foods, such as rice, grown in sites with high concentrations of mercury may also represent a source of organic and inorganic mercury exposure for some communities (Rothenberg, Windham-Myers, and Creswell 2014). Well-studied population groups vulnerable to mercury because of higher exposures are listed here. From the Global Mercury Assessment 2018 report, four populations of concern were identified based on existing datasets: 1) Arctic populations (mainly Inuit) who consume high-trophic fish and marine mammals; 2) tropical riverine communities (especially Amazonian) who consume fish, and in some cases may be exposed to mining operations; 3) coastal and/or small-island communities (including Indigenous Peoples) who rely substantially on seafood; and 4) individuals who either work or reside amongst ASGM sites. In addition to these relatively well-studied groups, other highly exposed groups for which there is awareness but relatively less data to draw firm conclusions include individuals living in mercury contaminated sites, certain occupational groups (e.g., chlor-alkali, dentistry), consumers of rice from contaminated sites, freshwater and marine fish consumers including sport fishers and Indigenous Peoples, and users of mercury-added products such as skin-lightening creams. In addition, there are certain ecosystems sensitive to mercury loading and methylation, and these may represent hotspots of biologically available methylmercury that warrant attention for those who consume local aquatic food items (see chapters 3 and 4). Coordinated studies that link human biomonitoring programs with data on environmental levels can help increase understanding of key processes that link mercury sources to human exposures.

5.6.2. Identification of sample population

Upon identifying a target population for investigation, the researchers would ideally sample all individuals from this target population, though achieving this is impractical (e.g., too many individuals to sample, it is prohibitively expensive, takes too much time and/or not everyone will agree to participate). Instead, researchers will sample a subset of the target population to realize a representative sample. Selection of the sample population needs to ensure that: 1) it is representative of the target population; and 2) there are sufficient number of people to yield valid information.

In order to select a sample population that is representative of the target population, it is necessary to understand the target population group’s socioeconomic and demographic profile. In addition, it is important to understand the target population’s mercury exposure profile (e.g., diet, occupation) and how this may change over time. The more specific the target population can be defined (e.g., age, sex, location, mercury exposure sources, seasonality, etc.), the easier it will be to identify a sample population with similar characteristics.
Figure 5.2. Population groups to consider. Within a country, exposures to mercury will be realized by all inhabitants (i.e. population universe), including members of the general population as well as members who are deemed vulnerable because of their lifestage or exposure situation. These population groups are not mutually exclusive as individuals may fall into multiple groups (e.g., those in ASGM sites may be exposed to both elemental mercury used in mining as well as methylmercury present in contaminated fish from local waterbodies). Once a specific target population is selected to focus upon (driven by their interests in consideration of the operational questions that support the Effectiveness Evaluation), steps need to be taken to help ensure that the sample population (yellow circle) is representative of the defined target population.

In order to select a sample population with a sufficient number of people, it is necessary to use statistical approaches that are aligned with the overarching aim of the biomonitoring effort. Guidance on statistical approaches is covered in relevant guidance documents from WHO (2018a), HBM4EU (2018b, 2018c, 2017b), along with many other resources (including online sample size calculators), and these need to be applied in a fit-for-purpose manner. To provide some additional context on possible sample sizes needed for a human biomonitoring study, the recent WHO guidance document on assessing prenatal exposures to mercury recommended a minimum of 250 pregnant women per site (WHO 2018a). In addition, the HBM4EU statistical plan (HBM4EU 2017b) mentions the need for at least 120 measures to derive a biomarker reference value in a defined population (based on guidance from the International Federation of Clinical Chemistry RefVal program). A scan of national biomonitoring programs covered in the Global Mercury Assessment 2018 biomonitoring dataset reveals average sample sizes in the several thousands of people (Basu et al. 2018). While statistical approaches can help ensure that there are sufficient number of people in the study to yield valid information, other considerations will factor into sample size decision making including the size of the underlying population, financial costs, trained personnel, infrastructure, timeframe, and spatial scale. Further, during the study design phase there should also be careful consideration of whether the population can be re-sampled in the future to permit temporal trends analysis.

The nature by which participants are recruited and studied should be carefully detailed following guidance from the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) initiative (Vandenbroucke et al. 2007). Ideally the sampling process is free from any biases, and participants are selected in a random manner. All studies should include a participant flow diagram to help explain the generalizability and validity of the results obtained from the sample population.

5.7. Human biomarkers

Human exposures can be assessed through the measurement of mercury concentrations in a number of different types of biological samples, and key approaches for mercury biomonitoring (including detailed protocols on how to take samples from study participants and perform analytical measurements of mercury in the laboratory) have been recently outlined by WHO (2018b) and HBM4EU (2018d, 2019).

The most commonly used and accepted biomarkers are measures of total mercury concentrations in hair, urine, blood, and cord blood, and their selection can depend on factors such as the potential source of exposure, chemical form, and exposure lifestage. These biomarkers, in particular, were the basis for the human exposure chapter in the Global Mercury Assessment 2018 report (Basu et al. 2018). Some elaboration on these accepted biomarkers is provided below.
Figure 5.3. Diagram of accepted mercury biomarkers (along the top) in correspondence with the different chemical forms of mercury that these biomarkers represent exposure to (along the bottom). Key population groups identified to be of concern from the Global Mercury Assessment 2018 are outlined in the middle of the figure, along with a horizontal band along the bottom that represents general populations.

5.7.1. Human hair

Analysis of hair for total mercury concentration is commonly used to assess exposure to methylmercury (which accounts for 80–90% of the hair’s total mercury content). Once incorporated, the mercury remains in the hair and this biomarker can therefore provide an integrated measurement of internal exposure to methylmercury. As hair grows at approximately 1 cm per month, exposures can be tracked over time by careful sampling (Lukina et al. 2021); for example, within person segmental hair analysis can integrate exposure data over several months, and examine differences across seasons or years.

Hair has the advantage that it is easy to collect, transport, and store, though in some communities there may be cultural objections to taking hair samples and in other groups (e.g., males, young children) short hair length may hinder proper sampling. Sampling should occur at the occipital region of the scalp for consistency and should be measured closest to the scalp to best reflect recent exposures (unless a longer temporal record is desired). In highly contaminated areas, there is a danger of external contamination of the hair, which can confound interpretation of the mercury measurement. For example, external contamination of hair by elemental mercury has been demonstrated in ASGM communities by use of mercury stable isotopes (Sherman et al. 2015). Therefore, when conducting studies in such contaminated sites care is needed in the interpretation of total mercury levels in hair. In such settings carefully analysing the hair for methylmercury, rather than total mercury, gives a better measure of dietary exposure especially when coupled with quality survey instruments, urine sampling, and biota measurements. Another potential challenge with hair monitoring in some communities may be the use of mercury-added cosmetic and beauty products. In such cases, hair total mercury levels may not accurately reflect dietary exposure to methylmercury. For this reason, when selecting individuals for hair monitoring, it is important to ascertain whether such mercury-added cosmetic products have been used. Further, when measuring hair mercury concentrations among such individuals, methylmercury (over total mercury) analysis is recommended.

5.7.2. Human urine

Analysis of urine for total mercury concentrations primarily provides information about recent (~1-2 months) exposure to inorganic and elemental mercury, although in people with high seafood consumption methylmercury may also contribute to the mercury content (Sherman et al. 2013). As the concentration of the analyte may depend on the dilution of the urine, which can vary, the measurement of mercury is often expressed in terms of its concentration per unit of creatinine or in relation to the
specific gravity of the urine sample. The collection of urine, as with hair, is relatively easy, non-invasive, and cost effective, and there are good protocols available from WHO (2018b) and HBM4EU (2018d, 2019).

### 5.7.3. Human blood

Mercury is measured in whole blood and this provides information about recent exposures (~1-2 months) to both methylmercury and inorganic mercury. Though many human biomonitoring programs focus on blood mercury measurements, the collection is invasive and the storage and transport of blood can pose certain logistical and financial barriers particularly in resource-limited settings.

In most population groups, the measurement of total mercury levels in whole blood is an accepted biomarker for methylmercury exposure as it correlates relatively well to seafood consumption (Sheehan et al. 2014). However, in certain population groups (e.g., those who do not consume much fish and seafood, or have relatively high exposures to inorganic and elemental mercury), total mercury may not be a good proxy for methylmercury exposure. Characterizing mercury chemical species or mercury stable isotopes in blood can provide an indication of potential sources, but these require careful sample preparation and advanced instrumentation. The US Centers for Disease Control and Prevention (CDC) now includes measures of blood methylmercury in NHANES and considers them more accurate in reflecting methylmercury exposures than measures of blood total mercury.

The measurement of total mercury levels in cord blood provides information about fetal exposure. Cord blood is collected following birth and often considered to be a non-invasive matrix, though this should be facilitated by a health care professional (e.g., nurse). Many jurisdictions have newborn screening programs in which newborn blood is sampled and archived as dried blood spots, and while mercury analysis of these dried blood spots shows promise they require careful consideration. Notably, dried blood spots are also collected in some demographic health surveys (e.g., USAID’s DHS Program) which are present in over 90 countries.

### 5.7.4. Integrated biomarker approach

Each biomarker can provide pertinent exposure information on the type of mercury (organic vs. inorganic) and timeline of exposure (recent vs. chronic). When multiple biomarker measurements are taken from a given individual (along with mercury speciation analysis and questionnaires), a deeper exposure assessment can be performed (i.e., under Tier 2 or Tier 3 biomonitoring activities).

Measurements of total mercury in hair and urine are particularly suitable (especially in resource limited settings) as they provide a relatively low-cost and non-invasive scheme to gauge exposure to the main forms of mercury. Further, with basic training, sampling and handling procedures are easy to implement, and quality assurance programs and suitable reference materials are also in place to help ensure comparability of measurement results (i.e., see good protocols from WHO (2018b) and HBM4EU (2018d, 2019) on how to take samples from study participants and perform analytical measurements of mercury in the laboratory). Biomarker measures can be further improved by also including survey instruments (see section 5.8) that collect pertinent information on the study population and exposure sources.

### 5.7.5. Biomarker measurements

A number of analytical methods (e.g., cold vapour atomic absorption spectrometry (CV-AAS) and cold vapour atomic fluorescence spectrometry (CV-AFS) are most widely used and accepted) are available to
 quantify the concentration of mercury in a given biomarker type, and these are detailed in a recent WHO guidance document (WHO 2018b) and by HBM4EU (2019). The selection of a particular analytical method will depend on factors such as availability of trained laboratory personnel and instrumentation. Regardless of the analytical method selected, it is important to practice careful quality control including the use of suitable reference materials (e.g., urine: INSPQ/Quebec; hair: NIES/Japan or IAEA/Austria; blood: NIST/US, INSPQ/Quebec) and attention to parameters such as detection limits, accuracy, and precision. It is also important to report the methods followed and QA procedures used. Analytical laboratories are encouraged to participate in quality assurance programs, such as the one run by AMAP/NCP, and these programs should be prepared to expand capabilities and provide assistance to nascent labs.

For the purposes of human biomonitoring (and as detailed above and in the included references), measures of total mercury content in a given biomarker will suffice in most cases. Such measures can be realized in under 10 minutes with minimal sample preparation using operationally simple, commercially available benchtop instruments that integrate sample decomposition with gold amalgamation and spectrophotometry.

5.8. Survey protocol

Combining the results of mercury biomarker measurements (section 5.7) with survey questionnaire information (e.g., sociodemographic data, occupational practices, dietary habits) from the same individual provides the basis for an assessment that can deepen understanding of exposure sources and routes as well as the extent, duration, frequency, and magnitude of exposure. Survey instruments relevant to mercury are available from WHO (Annex 3 in WHO (2018a)) and HBM4EU (2020b, 2020c, 2020a).

Surveys should be tailored for the target population (e.g., culturally appropriate, language, education level, relevant food items, lifestyle, and occupation) and have undergone proper pilot testing and validation. Those conducting surveys should have received training on proper methods to help ensure that valid and complete data are captured in a standard manner, and to identify and avoid possible sources of survey bias (for example, recall bias, estimations of serving sizes and frequencies). The survey data could also be amenable for capture into an electronic format.

5.8.1. Methylmercury exposures

Most populations worldwide are exposed to methylmercury through the consumption of fish and seafood (Sheehan et al. 2014; EFSA 2012). Thus, dietary intake of mercury from these items can be estimated if information is available on the: a) types and amounts (frequency and serving size) of food ingested per unit time (day or week); b) mercury concentrations in these food items (on a wet weight basis); and c) the participant’s body weight. Consumption of certain food items may vary seasonally, and mercury concentrations may vary across animal parts and be influenced by food preparation steps, and all of these need to be taken into account when conducting an exposure assessment. From a modelling perspective (chapter 6), it is also helpful to know the source of the food item (e.g., sampled locally or through international trade markets). As many of the food items that deliver mercury into human populations are also ones with high nutritional value, assessments should strive to examine risk-benefits (Mahaffey et al. 2011). Parties and relevant organizations could consider human biomonitoring efforts in geographic sites where biota are being sampled to maximize efficiencies and data quality (chapter 4).
Detailed protocols for developing dietary surveys are available from the UN/WHO (WHO 2008) and the US EPA (2016), and the HBM4EU has a comprehensive dietary questionnaire that may be adapted to fit particular needs (HBM4EU 2020a).

### 5.8.2. Elemental and inorganic mercury exposures

Human exposures to elemental and inorganic mercury may occur in occupational settings (e.g., in ASGM sites, chlor-alkali plants, and dentistry practices), from contact with certain products (e.g., dental amalgams, some skin-lightening creams, broken fluorescent bulbs and other waste products), and from environmental contamination (WHO 2008; Eagles-Smith et al. 2017; Ha et al. 2017; ATSDR 1999).

Identification of a target population based on these particular exposures should trigger the need to include screening level assessment surveys to deepen understanding of potential exposures. Examples of relevant screening level assessments for mercury are available from WHO (2018a) and HBM4EU (2020b, 2020c, 2020a). For the ASGM sector, guidance from WHO provides templates and tools for conducting assessments to provide an evidence base for the development of public health strategies required for National Action Plans (WHO 2021b). There is also a survey from a UNIDO/UNDP/GEF-sponsored initiative that is often used (Veiga and Baker 2003), which needs to be applied with careful attention to tease apart different job tasks, the proximity of ASGM sites to households, and location of smelting and ore processing sites. For dentistry, a collaboration between the American Dental Association and academics yielded a survey tool to relate occupational practices with exposure biomarkers (Goodrich et al. 2016), and the HBM4EU has a survey with pertinent questions concerning personal amalgams (HBM4EU 2020a).

### 5.9. Management and analysis of human biomonitoring data

#### 5.9.1. Existing and future data

Existing data, as contained initially in the WHO-sponsored, Global Mercury Assessment 2018 human biomonitoring dataset (Basu et al. 2018), can be updated to help establish a “baseline” for human biomonitoring under the Effectiveness Evaluation. In terms of future data, we can expect, with very high confidence, that biomonitoring data will be available from government-led national biomonitoring programs as well as academic-led cross-sectional and birth cohort studies. In addition, to help fill data gaps in a coordinated manner and build capacity, Parties and relevant organizations are encouraged to consider recent guidance from the WHO on a harmonized approach for conducting new biomonitoring activities (WHO 2018a).

#### 5.9.2. Data quality

Quality practices are necessary to help ensure that biomonitoring results are valid, free of bias, and comparable across studies and regions. In terms of ensuring that field work is conducted properly, information presented earlier under Sections 7.5 (Populations) and 7.7 (Surveys) should be consulted, along with resource documents from the WHO (2008, 2018a) and the STROBE initiative (Vandenbroucke et al. 2007). It is essential that studies collect critical details on the sample population (e.g., age, sex, location, sample month/year), how they were recruited, and details on sources and routes of mercury exposure. In terms of biomarker measures, information presented earlier under section 5.7 (Human biomarkers) should be consulted so that studies use proper reference materials, participate in inter-lab
comparison programs, and report on analytical parameters such as detection limits, accuracy, and precision.

Based on guidance from the US National Toxicology Program’s (NTP) Office of Health Assessment and Translation (OHAT 2015), and as considered as part of the Global Mercury Assessment 2018 biomonitoring dataset, a Risk of Bias score can be derived for each study that considers: a) participant selection bias (e.g., selection method, demographics, exposure characteristics, timing of recruitment); b) exposure detection bias (e.g., quality of the methods used to measure the mercury biomarkers, recall bias); and c) statistical and other bias (e.g., biomarker distribution, reporting mercury exposure sources). Such a score can help give users of the data a frank assessment of its quality, and be used to flag potential concerns.

5.9.3. Data exchange

Paragraph 1 (d) of article 17 of the Convention calls for Parties to facilitate the exchange of epidemiological information, in close cooperation with the WHO and other relevant organizations, as appropriate. To facilitate the implementation of that article of the Convention, cooperation for the compilation and exchange of data via a centralized database may be considered.

For each biomonitoring study to be included in such a database there is a need for minimal essential information to help ensure that studies can be compared. These include group-level data on: sample population characteristics (population type, sample size, age, sex, education, socioeconomic status, personal amalgams, city/region/country, day/month/year), analytical measurements (sample size, biomarker type, speciation information, quality control including detection limit, accuracy, precision, and use of reference materials), and mercury values (count (n)), percentiles including 10th, 25th, 50th, 75th, 90th, and 95th values; additional measures of central tendency (variance) including mean (SD) and geometric mean (95% CI); indication of data normality; these align with guidance from the HBM4EU statistical analysis plan (HBM4EU 2017b)). Strategies for dealing with missing data and measures below detection limits are provided in the HBM4EU statistical analysis plan (HBM4EU 2017b). Key information from surveys (e.g., dietary intake values; occupational practices; other exposure sources) needs to also be extracted and summarized. The data should be aggregated for the entire sample population as a primary level summary, as well as for key sub-groups (e.g., different lifestages, sexes, locations, occupational categories) as part of a secondary level summary. Finally, studies must name the ethics board that approved their work. Section 11.2.2 of the HBM4EU statistical analysis plan provides a good list of variables specific to mercury organized into exposure levels, time trends, geographical comparisons, and exposure determinants that largely align with the information listed here (HBM4EU 2017b).

The focus of the human biomonitoring data should be on a population group. While compiling individual-level data may permit deeper scientific analysis, realizing this for research of human subjects is extremely challenging owing to ethical, privacy, logistical, and other concerns. The WHO guidance document provides guidance on handling individual-level data i.e. participating countries conduct statistical analysis in-country, and then submit anonymized summarized data to a central database for international-level analyses (WHO 2018a). A similar approach may be taken for group-level data as well, with good details offered by HBM4EU on handling both individual- and group-level data (HBM4EU 2017b).
5.9.4. Data analysis

Statistical analysis of human biomonitoring data may help address the questions that support the Effectiveness Evaluation (table 2.1). Detailed guidance on statistical analysis of human biomonitoring data is offered by HBM4EU, and it covers aspects such as treating missing data, time trends analysis, geographic comparisons, and uncertainty analysis (HBM4EU 2017b). Five key statistical analyses are listed below that align with the overarching policy-relevant questions. More sophisticated aspects of data analysis (especially modelling) are provided in chapter 6, and here basic guidance is provided on how to analyze mercury human biomonitoring data.

Descriptive statistics: Descriptive statistics should be used to summarize key features of the sample population and their exposures to mercury. This information can be used, for example, to characterize spatial variability, and help identify hotspots and exposure sources. The data can also be used to indicate the percentage of those sampled with mercury biomarker values that exceed a guideline value or reference range at a certain place and point in time (these are summarized in Basu et al. (2018). Such descriptive information can then be represented visually on a map with a color scale as done for an assessment of human biomarker values from across Europe (Višnjevec, Kocman, and Horvat 2014).

Exposure assessment: To increase understanding of possible sources and routes of mercury exposure, regression-based approaches may help associate mercury biomarker measures (dependent variable) with independent variables drawn from the survey data (e.g., dietary intake, occupational practices). There are many published studies of this kind for a diverse range of mercury exposure scenarios.

Temporal analysis: Over time changes can be gleaned if repeated monitoring is performed in the same population over time. This requires that the geographic scale (local to national to global) and the target population (e.g., background, specific vulnerable, life stage, etc.) be defined, and then differences in mercury biomarker measures be compared. Depending on the context, seasonality of sampling may be an important consideration here. Section 6 of the HBM4EU statistical analysis plan provides detailed guidance on temporal trends analysis (HBM4EU 2017b).

Attributive analysis: If temporal changes in mercury biomarker levels are found, stakeholders will want to know if changes are attributed to actions taken under the Minamata Convention. This will require exposure assessments and temporal analysis to be combined, and with consideration of discrete policy actions taken. Successful examples are, as discussed in the Global Mercury Assessment 2018 report: a) decreasing blood and hair mercury levels have been reported in population groups from the United States, Denmark, the Faroe Islands, and several Arctic communities that may be linked with dietary consumption advisories and/or changing dietary habits; and b) decreasing urinary mercury levels among the general US population, German children, and some dental professionals is likely associated with the development of encapsulated amalgams, the increasing use of composite resins, and the overall awareness of occupational and environmental risks associated with mercury use.

Risk assessment: One of the ultimate goals of the Minamata Convention is to protect human health from mercury. Established risk assessment frameworks (e.g., EFSA 2012) may be used to calculate the nature and probability of mercury-associated adverse human health effects. From such data, burden of disease estimates and economic costs may be calculated, and changes over time may be explored under actual conditions and future scenarios using modelling tools.
5.10. Communication

Communication of results is a critical aspect of human biomonitoring. The HBM4EU program offers guidance on how human biomonitoring data could be organized into a report (HBM4EU 2020d), and the WHO offers guidance on how researchers should engage with stakeholders throughout the project’s life course, and how biomonitoring findings should be shared with study participants, the general public, public health professionals and policy makers (WHO 2018a). In addition, particular consideration is needed with regards to contaminant research pertinent to Indigenous populations, where a partnership approach and equitable engagement ensures successful communication of monitoring and research results (see AMAP for examples of positive and negative experiences (e.g. AMAP 2021). Parties and relevant organizations may also decide on if (and how) the data is used for risk management.

5.11. Conclusions

Human biomonitoring data can help address the operational questions that will support the Effectiveness Evaluation (see Table 2.1). The information in this chapter provides essential guidance (and links to key resources) for Parties and relevant organizations to consider in terms of using existing, and generating new, human biomonitoring data for the Effectiveness Evaluation.

In terms of using existing biomonitoring data, several databases and resources exist, and these can be used to help understand human exposures to mercury before the Minamata Convention’s entry into force (i.e. help establish the baseline).

In terms of data to be realized during the Effectiveness Evaluation period, there are two sources to consider. First, biomonitoring data in the future are expected to be realized from existing government-led national biomonitoring programs, regional initiatives, and/or academic-led studies. Second, Parties and relevant organizations can further support the Effectiveness Evaluation by implementing new biomonitoring studies in a harmonized way so that they are purposefully designed to fill data gaps, and build capacity.

Human biomonitoring data can be designed as part of a Tiered approach to inform new monitoring programmes or improve existing ones (see section 5.4 and the annex to this document). Briefly, Tier 1 is for those seeking to create a human biomonitoring programme, or expand a minimal programme, but that may not have sufficient resources to implement the actions in Tier 2. The goal of Tier 1 should be to focus on a vulnerable sub-population (section 5.6) and take total mercury measurements in blood, urine, or hair (section 5.7). This activity should ideally be repeated in the same population every 2-5 years. A good starting point for Tier 1 is the recent guidance from the WHO to characterize prenatal mercury exposure (WHO 2018a). Tier 2 aims to realize information that will help address all operational questions in Table 2.1, and thus calls for more in-depth analysis of the Tier 1 sub-population group, or incorporation of mercury biomonitoring into other, in-depth health surveys or cohort studies. Tier 3 aims to increase understanding of key processes that link mercury sources to human exposures, and thus resource-intensive research methods and approaches are required.

There are essential elements to all human biomonitoring studies that need to be considered, and these are outlined in Figure 5.1 and elaborated upon in this chapter. Key elements include: a) defining the target and sample population (which usually focus on groups vulnerable to mercury i.e. early lifestages or those with relatively high exposures); b) selecting and measuring the appropriate biomarkers to help tease apart exposure to different sources and forms of mercury (with total mercury measurements in
hair, urine, blood and cord blood being most commonly used and accepted); c) administering surveys to
gather supportive information (e.g., on socio-demographics, occupational practices, dietary habits) to
depth; and d) managing and analyzing data as per the guiding policy question. All these
aspects must be performed in a responsible and ethical manner.
Chapter 6. Cross-media data management, modelling and analysis

6.1. Introduction

Chapters 3, 4 and 5 provide guidance on the collection, management and analysis of data in air, biota and from human biomonitoring. By analysing monitoring data, temporal and spatial trends in the levels of mercury in specific environmental media or human matrices can be derived with confidence intervals. These trends provide a first-level indication of whether the Convention may be contributing to protecting human health and the environment from the adverse effects of mercury. Analyses of the monitoring data collected in each medium separately will be highly informative, and cross-media analysis incorporating the known mechanistic connections between media can provide further information, adding to the scientific weight of evidence that can inform the Effectiveness Evaluation. This chapter elaborates on how these monitoring data can be used in an integrated manner, where combining multiple complementary analysis approaches to answer the same question will improve robustness. This will facilitate understanding of the spatial and temporal trends and patterns of mercury observed in the environment and humans, and the impact of actions motivated by the Convention.

Because the connections between monitoring media are not necessarily direct and instantaneous but do depend largely on known or suspected physical processes, mechanistic models explicitly representing these processes are a valuable tool for interpretation of monitoring results and can thereby contribute to the Effectiveness Evaluation. However, as the complexity of the modelled system increases, identifying all the relevant processes and quantifying them correctly becomes more challenging. In such cases, mechanistic models can be supplemented with different kinds of statistical models. Attribution of observed trends to specific drivers such as direct anthropogenic mercury releases, legacy mercury, natural process-driven releases, and non-mercury environmental or behavioural drivers requires the use of models which resolve the intervening processes supplemented or calibrated by empirical statistical approaches. From primary release to human exposure, mercury can undergo many physical and (bio-)chemical changes which interact with each other over a large range of timescales and can be influenced by human behaviour. This makes cross-media analysis involving both mechanistic and statistical modelling in all relevant media an important part of the weight of evidence useful to evaluate effectiveness of the Convention. Moreover, evaluating the effectiveness of the Convention requires separating the impacts attributable to the Convention from changes that occur due to other factors, such as economic growth, climate change, etc. While monitoring data shows the impact of all of these factors, modelling can help attribute the changes to the different drivers.

Monitoring data and other ancillary observational data can be used in a variety of ways in concert with mechanistic and statistical models to quantify the effectiveness of Convention measures. Data from each medium can be used to evaluate that medium’s model representations, and to identify situations where a given model is or is not appropriate for use. Monitoring data from one medium can also be used as input to models to explicitly connect outcomes in that medium to outcomes in other media (e.g. wet deposition can serve as input to an aquatic ecosystem model to estimate fish concentrations), or to models which can attribute those trends to specific sources or drivers. Tables 6.1 and 6.2 summarize, for monitoring data and ancillary observational data respectively, the data, metadata, and other information that can facilitate cross-media analysis and modelling. Where available, monitoring data from ocean and freshwater, although not core media in this guidance, may provide important information to strengthen the accuracy of analysis and prediction.
The analyses discussed below fall into two main categories of approach. The first is the top-down approach, which directly uses monitoring data and statistical relationships to relevant variables to infer importance of specific drivers from the observational data. The second is the bottom-up approach, which uses mechanistic models representing physical processes to produce estimates of the quantities that are observed based on inputs to the modelled system. These two approaches can be interpreted as propagating information in opposite directions, the former from observed quantities to their drivers and the latter from the drivers to observable quantities. Both approaches can be useful and are discussed further in the following sections.

Table 6.1. Information from monitoring data. Listed for each medium and tier are the primary monitoring data, metadata, ancillary data for interpretation and to aid in analyses, and the analyses for which those data can be used.

<table>
<thead>
<tr>
<th>Monitoring category</th>
<th>Observation Data</th>
<th>Metadata</th>
<th>Ancillary Data</th>
<th>Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air - Tier 1</td>
<td>Total or gaseous elemental mercury levels; measurement/method uncertainty; Wet deposition</td>
<td>Latitude; longitude; elevation; Sampling time, frequency, duration; averaging methods; sampling method</td>
<td>Proximity to known point sources; type (urban/regional/background); meteorological variables; measurement/method uncertainty</td>
<td>Temporal trends ● Atmospheric model evaluation (for GEM) ● Spatial variations ● Input for local-scale modelling ● Back-trajectory analysis ● Bottom-up attribution analysis15</td>
</tr>
<tr>
<td>Air - Tier 2</td>
<td>Air - Tier 1 and Speciated reactive mercury; high-resolution PBM and GOM; Dry deposition of mercury; mercury throughfall</td>
<td>Air - Tier 1</td>
<td>Air - Tier 1 and deposition of Sulfate; Land Cover; Land Use; Leaf Area Index; Air Quality Tracers (e.g., SO2, CO2, CO, PM2.5, O3)</td>
<td>Air - Tier 1 and ● Estimate air-ocean and air-terrestrial mercury exchange ● Covariate profiling ● Top-down attribution analysis ● Speciated and isotopic measurements</td>
</tr>
<tr>
<td>Air - Tier 3</td>
<td>Air - Tier 2 and mercury isotopes; additional speciation measurements</td>
<td>Air - Tier 1</td>
<td>Air - Tier 2</td>
<td>Air - Tier 2 and ● Combined “top-down” and “bottom-up” attribution analyses ● isotopic fingerprinting</td>
</tr>
</tbody>
</table>

15 The term "Bottom-up" is being used to refer to a process-based analysis estimating effects of drivers on observable quantities. The term "Top-down" is being used to refer to an observation-based analysis for identification/estimation of drivers.
| Biota - Tier 1 | Tissue/organ mercury and/or methylmercury levels; measurement/method uncertainty; distribution statistics or quantiles | Geolocation or water body name; Spatial coverage; sampling time period; method info; tissue/organ type; habitat; wet or dry weight | Population sample size; species; length/mass; trophic position/diet info; age; sex; maturity stage; carbon and nitrogen isotopic data; lake size; known point source or sediment contamination; water temperature, DOC, pH, TSS, nutrient concentrations | ● Spatial variations  
● Temporal trends  
● Input for local exposure modelling |
|---|---|---|---|---|
| Biota - Tier 2 | Biota - Tier 1 | Biota - Tier 1 | Biota - Tier 1 and carbon and nitrogen stable isotopes; water DOM/DOC/TOC, TSS, salinity, DO, (pH). N and P, Chl-a; total mercury in sediment; GEM in air; wet deposition; meteorological data | Biota - Tier 1 and  
● “Top-down” biota mercury attribution  
● Food web model evaluation |
| Biota - Tier 3 | Biota - Tier 1 | Biota - Tier 1 | Biota - Tier 2 and mercury stable isotopes in biota and suspected source-matrices; chemical tracers related to known drivers; diet information; stable isotopes of prey organisms; food web structure | Biota - Tier 2 and  
● Combined “top-down” and “bottom-up” biota mercury attribution  
● Isotopic fingerprinting |
| Human - Tier 1 | Total mercury levels in hair, blood, or urine ($10^{th}$, $25^{th}$, $50^{th}$, $75^{th}$, $90^{th}$, and $95^{th}$ percentiles); | Geolocation or city/country/region; Population sample size; Spatial coverage; population type; sampling time period; method info; type of biomonitoring sample; ethics board diet info; age; sex; known occupational and other exposures; education, socioeconomic status, amalgam status; additional measures of central tendency (variance) including mean (SD) and | ● Spatial variations  
● Temporal trends  
● Exposure model evaluation  
● Input for local health impact / risk assessment modelling  
● Guideline value exceedance statistics  
● “Top-down” |
<table>
<thead>
<tr>
<th>Medium</th>
<th>Data</th>
<th>Metadata</th>
<th>Ancillary Data</th>
<th>Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soils</td>
<td>Mercury levels; measurement/method uncertainty; surface fluxes</td>
<td>Latitude; longitude; depth; Sampling date; averaging methods</td>
<td>Presence of known point sources; soil horizon; land use; carbon concentrations</td>
<td>• Terrestrial model evaluation&lt;br&gt;• Input for local-scale modelling&lt;br&gt;• Input for atmospheric modelling&lt;br&gt;• Atmospheric model evaluation</td>
</tr>
<tr>
<td>Vegetation</td>
<td>Mercury levels; measurement/method uncertainty; exchange fluxes; litterfall fluxes</td>
<td>latitude; longitude; Sampling time; averaging methods</td>
<td>vegetation type; NDVI; carbon fluxes</td>
<td>• Terrestrial model evaluation&lt;br&gt;• Input for local-scale modelling&lt;br&gt;• Input for atmospheric modelling&lt;br&gt;• Atmospheric model evaluation</td>
</tr>
</tbody>
</table>
### Food items and other products

<table>
<thead>
<tr>
<th>Food items and other products</th>
<th>Methylmercury and total mercury levels; statistical distribution information</th>
<th>name of product; type of food item; country/region; Sampling time period</th>
<th>Consumer population; exposure type (diet, skin, etc.)</th>
<th>● Input for exposure modelling</th>
</tr>
</thead>
</table>

### Freshwater

<table>
<thead>
<tr>
<th>Freshwater</th>
<th>Mercury levels; methylmercury levels; measurement/method uncertainty</th>
<th>latitude; longitude; depth; Sampling time; averaging methods; water body name</th>
<th>dissolved and particulate carbon concentrations; temperature</th>
<th>● Input for food web modelling</th>
</tr>
</thead>
</table>

### Ocean

<table>
<thead>
<tr>
<th>Ocean</th>
<th>Mercury levels; methylmercury levels; measurement/method uncertainty</th>
<th>latitude; longitude; depth; Sampling time; averaging methods; water mass name</th>
<th>dissolved and particulate carbon concentrations; nutrient concentrations; temperature; salinity; dissolved oxygen</th>
<th>● Ocean model evaluation</th>
<th>● Input for food web modelling</th>
<th>● Input for atmospheric modelling</th>
<th>● Atmospheric model evaluation</th>
</tr>
</thead>
</table>

### Sediment

<table>
<thead>
<tr>
<th>Sediment</th>
<th>Mercury levels; methylmercury levels; mercury accumulation rates; measurement/method uncertainty</th>
<th>type of sediment; latitude; longitude; water depth; sediment depth; dating info; dating method</th>
<th>accumulation rates; total organic carbon; grain size</th>
<th>● Input for watershed modelling</th>
<th>● Input for food web modelling</th>
<th>● Mass balance model evaluation</th>
</tr>
</thead>
</table>

### Snowpack

| Snowpack | Mercury levels; exchange fluxes; measurement/method uncertainty | latitude; longitude; depth; Sampling time; averaging methods; sampling methods | snow depth; accumulation rates; snow density | ● Atmospheric model evaluation | ● Input for local-scale modelling |

### 6.2. Maximizing scientific weight of evidence

Some media-specific analyses are outlined in chapters 3, 4, and 5, and can be useful tools to inform Effectiveness Evaluation via single-medium monitoring data. Chapter 3 discusses management, analysis and evaluation of atmospheric mercury data and provides tools for obtaining a more holistic picture of the state of mercury in air by adding value to the monitoring data that is collected. Chapter 4 enumerates the primary and ancillary monitoring measurements for biota which can be used for time series analysis accounting for variability associated with multiple factors, and discusses ecosystem sensitivity analysis to identify and prioritize sites for most effective use of limited monitoring resources.
Chapter 5 highlights descriptive statistics and temporal analysis on human biomonitoring data used to summarize population exposures to mercury and how they change in time, exposure assessments using survey data to associate biomarker measurements to possible sources, and risk assessment to connect to human health. Such tools can also be combined into integrated analyses across media to provide further information. Chapter 2 presents operational questions (Table 2.1) which serve as a guide to producing a continuum of evidence, and the following sections describe tools and analyses which can be used to maximize the scientific weight of evidence used to evaluate the effectiveness of the Convention. Furthermore, the Supplementary Material elaborates on laboratory intercomparisons for identifying biases and uncertainties in air monitoring, time trend identification, covariate analysis for source identification, and backwards trajectory models for source-receptor relationships.

6.2.1. Estimation of background and impacted levels of mercury

Analysis of monitoring data from sites chosen and categorized to represent background and impacted locations can directly estimate levels of mercury for these types of areas. Aggregating results of total air mercury and wet deposition monitoring by site type (e.g. urban/point-source-influenced/background) can show high-level source influences using data from all tiers. Similarly, biota monitoring data of all tiers at locations identified as background or affected by anthropogenic sources can be used to estimate these biota-specific levels of mercury. Summary statistics comparing some subpopulations from human biomonitoring can be used to establish mercury levels for some background and impacted locations where representative subpopulations have been sampled. These can include both mercury levels and guideline value-exceedance statistics. These most basic analyses can usually be done with a high degree of confidence using basic monitoring strategies, yet can provide valuable information for the effectiveness evaluation.

6.2.2. Identification of trends over time

The effective identification of temporal trends should yield key pieces of information: the magnitude of the trend, an associated confidence interval, and a summary measure of the statistical significance of the trend. It is more difficult to identify trends in areas with high temporal variability because the magnitude of the trend is more likely to come with a relatively wider confidence interval and lower statistical significance. Including the confidence interval and statistical significance helps to avoid over-interpretation of observed trends.

Statistical analyses can be performed on time series data from air monitoring sites (Tier 1, 2, and 3) to identify observed trends which take into account sources of temporal variability. The significance of an upward or downward trend across a time period and its standard error can be identified using appropriate statistical tests. The magnitude of the trend and its confidence interval can be obtained using this information and statistical methods which are robust to outliers and data which deviate from linear time behaviour.

For biota monitoring (Tier 1, 2, and 3), if individual sample data are available, generalized linear modelling can account for time-variations in mercury levels in a way that controls for drivers included in the measured ancillary data and metadata. Consistency in data quantities available across monitoring locations is important for accurate application of this method.

Human biomonitoring (Tier 1, 2, and 3) differences across time can be identified on a subpopulation basis if repeated monitoring is performed in the same population over time. This requires that the
spatial coverage, sampling timing and population be well defined, to be able to control for covariates such as seasonality, life-stage, and human activities (e.g., diet, occupation), etc.

Bottom-up modelling can be performed to quantify expected trends or relative trends in locations, forms of mercury, and matrices that lack direct monitoring. Spatially resolved models which show consistency with observed trends in monitoring locations can estimate expected trends in other locations. Site-specific modelling can extend the observed trends in monitoring media to other media and to exposure and health impacts through cause-effect relationships.

Modelled temporal variability is a quantity which requires careful consideration of model inputs and assumptions. Variability driven by environmental variables such as temperature and weather will only be quantifiable by a given model on the timescale that those variables are represented in the model. Often input variables are averaged over days to years, depending on the model and the input, and therefore shorter-timescale variability will be under-represented in model output.

**Model Description 6.1: Atmospheric models**

Atmospheric chemistry transport models represent the fate of mercury upon release to the atmosphere. They represent the chemical and physical changes in the form of the released mercury using experimentally or theoretically determined reaction rate and partitioning coefficients. Atmospheric models can be global- or regional-scale gridded models or trajectory models which trace the dispersion of air parcels forward from sources or backwards from receptors. To trace emissions to receptors, they require specification of the magnitude and spatial distribution of releases of mercury to air: as anthropogenic and geogenic direct emissions, as well as terrestrial and ocean fluxes of legacy mercury. Since these models directly simulate atmospheric concentrations and deposition, measurements of these quantities are best suited for evaluation. In comparing these quantities, it is important to consider that concentration and deposition measurements are often performed at a single point, while gridded atmospheric models represent the values over some area depending on the model grid size, and gridded and trajectory models rely on the resolution of the underlying meteorological data. Therefore, these models can be limited in their ability to resolve high local- or small-scale variability, even if their ability to do so can be improved with smaller grid size and higher observation density. The averaging time and sampling frequency of the measurements compared to those of the model output should also be considered. The relevant timescales for large-scale changes in atmospheric mercury are months to years. For simulation of trends, atmospheric models must be driven using time-varying inputs of both anthropogenic and legacy mercury to the atmosphere.

- **Strengths:** Bottom-up source attribution, large-scale spatial variability
- **Weaknesses:** Reliance on accuracy, temporal coverage, and availability of emission inventories
- **Readiness:** Multiple available models

**Model Description 6.2: Ocean models**

Global ocean models represent the marine fate of mercury deposited to the oceans from the atmosphere and entering the oceans via rivers. They require specification of the magnitude and spatial distribution of wet and dry deposition as well as river concentrations as inputs. These models simulate transport by ocean currents, mercury methylation, particle partitioning and sinking. Since ocean models directly calculate total seawater mercury and methylmercury concentrations,
observations of these quantities are most comparable. In these comparisons, important considerations are comparing a near-instantaneous measurement with a longer time-averaged model value and comparing point measurements against model values representing a large area. Coastal and heavily river-influenced areas will be more sensitive to local releases via river inputs, while open-ocean measurements will be more sensitive to atmospheric inputs. The relevant timescales for ocean mercury are years to centuries. Simulation of trends of ocean mercury concentrations will require that inputs of riverine mercury releases as well as atmospheric deposition to ocean be time-varying.

- Strengths: air-sea exchange impacts, decadal time-scale changes
- Weaknesses: Propagates uncertainties in inputs from atmospheric models
- Readiness: Multiple available models

Model Description 6.3: Terrestrial models

Terrestrial models represent the exchange of mercury between vegetation and soil reservoirs via processes associated with biochemical transformations of carbon by plants and soil microbes. These models use atmospheric mercury concentrations and deposition as inputs to calculate the plant uptake, throughfall, litterfall, and soil uptake of mercury, as well as soil evasion fluxes of mercury due to the microbial breakdown of mercury-containing carbon compounds in soils. The breakdown of mercury-containing compounds takes place over a wide range of time scales, meaning that terrestrial models account for mercury responses to changes ranging from seasonal to over centuries and longer. These models are useful for estimating legacy contributions to environmental mercury levels.

- Strengths: Legacy and environmental driver source attribution, long-time-scale influence
- Weaknesses: Reliance on historical input information, large amount of ancillary data required, lack of data on terrestrial ecosystems
- Readiness: Emerging applications for multi-media model coupling

6.2.3. Characterization of representative levels and spatial patterns

Total gaseous and elemental mercury concentrations from comparable measurements can be compared across monitoring sites and between types of monitoring sites to quantify spatial patterns in air mercury levels. Mapping these concentrations can show geographic patterns. Similarly, wet deposition measurements can be compared across sites and between site types. This type of analysis can be performed for measurement sites of all tiers.

Spatially resolved atmospheric models can estimate the level and form of mercury across a wide range of locations and times, including at locations and times not directly covered by monitoring. This model output can be used to supplement monitoring findings by filling the gaps between monitoring locations. These models can also estimate how representative an observation is by quantifying the expected spatial and temporal variability in the observation's vicinity. By quantifying the representativeness of an observation, the models can elevate its evaluative power. For example, in regions where spatial gradients are expected to be small according to models, a single observation site can effectively monitor a wide region. This means that models can also be used to inform monitoring locations, suggesting
denser monitoring in more spatially variable areas. In regions and locations with significant local
sources, smaller-scale modelling could be significant.

Models can be used to extend the observed spatial patterns of mercury in one observed form or matrix
to other forms or matrices, because they can take inputs consistent with the observations in one
medium and simulate the resulting patterns in the medium/media they represent.

Subpopulation summary statistics from biomonitoring can be used to establish baseline mercury levels
and potentially broad spatial patterns depending on subpopulation locations. These can include both
mercury levels and guideline value-exceedance statistics. Comparison could be done across identified
vulnerable subpopulations (with Tier 1 monitoring), or across national or other subpopulations (with
Tier 2 or 3 monitoring).

A standardized method of visualization and summary analysis would facilitate communication of
combined monitoring and modelling findings. The most common form of visual comparison for spatially
resolved models with collections of observational data is a coloured map of modelled values with the
corresponding observations of the same quantity overlaid as coloured dots using the same colour scale.
A standardized choice of colour scale and map projection would aid visual comparison between
different models of the same type. An indication of the underlying model resolution in the form of a grid
can aid visual interpretation of spatial variability. Colour maps should be chosen with consideration of
viewers with colour vision deficiency, and be diverging for quantities that can be positive or negative.
Overlaid hatching and special symbols can be used to annotate whether mapped trends are statistically
significant.

Interactive web-based tools to support model data exploration and access could increase the reusability
of model output. In the case of spatially resolved models, an online platform which allows for data
selection and visualization, with subsetting by medium, quantity of interest, location(s), etc. would allow
for maximum reuse by users for the purpose of smaller spatial-scale analysis.

Model Description 6.4: Generalized linear/additive models (GLM/GAM)

A generalized linear model (GLM) is a generalized version of linear regression which does not assume
that the response variable error is normally distributed and does not assume that the response
variable changes linearly with changing predictors. This added flexibility allows a GLM more
explanatory power for quantities such as mercury concentrations in monitoring media which can have
complicated responses to specific observable drivers. GLM can be further extended to generalized
linear mixed models (GLMM), in which predictor variables additionally contain a random component
to their effects, and generalized additive models (GAM), in which predictor variable coefficients are
generalized to functions.

These types of models can be used with monitoring data to control for and attribute observed
variability to specific independent variables. These observation-driven relationships to drivers of
variability can be used as a “top-down” constraint for attribution. Because the valid range of these
models is determined by their training data, monitoring data should share common comparable
ancillary data across sites to most effectively implement this type of analysis. Separate training and
testing data subsets should be used to avoid overfitting, and model assumptions should be checked
by examining residuals.

- Strengths: Top-down attribution, application not specific to any given medium, monitoring-
driven
• Weaknesses: Reliance on wide-ranging comparable data and ancillary data
• Readiness: Widely-used methodology

6.2.4. Estimation of information on source attribution

Models can not only calculate mercury levels and trends, but can also quantify the contributions to those values by specific drivers. Because emissions sources are direct inputs to atmospheric models, these models can be used to isolate emissions responses in observed and modelled trends in mercury concentrations and wet deposition (Tier 1) using a bottom-up approach. Such models can therefore be used along with observed trends to quantitatively attribute the trends to specific source types. This is true of types of sources, such as primary anthropogenic vs. legacy, as well as the relative importance of local sources vs. global sources using models that can resolve these types of sources individually.

Where ancillary air information is also available (Tier 2 and 3), a top-down approach can attribute observed trends to sources and drivers. At these locations, combining the monitoring-driven top-down approaches with bottom-up attribution from atmospheric models can balance explanatory and predictive power to provide more robust attribution estimates than either method individually.

Biota monitoring and ancillary data (Tier 2 and 3) can be used in top-down modelling to estimate the contributions of different sources and large-scale drivers to biota mercury levels and trends. These sources and drivers can be further attributed by bottom-up modelling to different types of sources using a combination of watershed, mass balance, atmospheric, and/or food web models. The number of models/media required to attribute mercury levels and trends will vary from site to site and depend on the relative contributions of drivers in each medium. At intensive monitoring locations (Tier 3), top-down and bottom-up approaches can be combined to “calibrate” mechanistic model input parameters.

Quantifying the contribution of sources of natural and legacy mercury requires some level of multimedia approach. For single-medium models, inputs corresponding to these types of sources can be varied in a way that reflects the changes occurring in the source media. Coupled-media models can directly simulate the concurrent changes in legacy fluxes between media in a self-consistent manner while changing only primary releases as model inputs. For site-specific modelling, multimedia mass balance models present a tool for attribution that includes legacy sources.

The attribution of trends and changes in mercury levels in all monitoring media to environmental drivers unrelated to the Convention can also be performed using a top-down approach where the necessary ancillary data is available (Tier 2 and 3 sites), and can in some cases be performed using a bottom-up approach where those drivers can be explicitly changed in model scenarios. In the atmosphere, weather patterns and climate cycles can lead to variability in mercury levels and deposition through changes in temperature and precipitation. Atmospheric variability can translate to the surface ocean, and the ocean has analogous climate cycles that can affect observed trends. Terrestrial systems are strongly affected by land-use changes, and changes in the cryosphere can propagate effects to the atmosphere, aquatic and terrestrial environments. In biota, variability in temperatures and food web structures as through prey availability can cause changes in biota mercury levels unrelated to anthropogenic mercury emissions. These changes are unrelated to the Convention but can have impacts on observed trends, and quantifying their contribution allows a more accurate evaluation of Convention effectiveness.

Variability in environmental drivers are especially relevant to site-specific and small spatial scale trends.
Changes in human biomarker levels can be attributed to drivers through exposure assessments using in-depth survey data and sophisticated biomarker analyses that include, for example, multiple biomarkers, mercury speciation analysis, and/or mercury stable isotopes (Tier 2 and 3). A top-down approach can identify contributions from changing dietary habits, occupational and other exposures that can be estimated through the survey based on Tier 1 information. Attributions to measures influenced by the Convention can already be made at this level when considering the drivers for the changed behaviour in a careful and scientifically sound manner. When further adding ancillary monitoring and other information from Tier 2 and 3, including known dietary intake quantities due to biota mercury levels, even smaller responses can be attributed to behavioural changes influenced by the Convention or changing mercury concentrations in the diet.

When adding bottom-up modelling to the above-described approach, improved explanatory power that includes even more factors influenced by the Convention can be obtained. This comprehensive long-term goal of the Effectiveness Evaluation will require an accumulation of monitoring data and analyses to provide information with a high degree of confidence, but most of the attribution analysis steps will be able to provide useful information immediately based on Tier 1 data.

**Model predictive and explanatory power**

Mechanistic models share an overall structure whereby they are designed to simulate or represent a collection of interactive physical/biochemical processes involving mercury in one or more media and forms, and require inputs representing the flow of mercury into the scope or domain of the model as well as biogeochemical and physical environmental conditions. The mathematical representation of physical/chemical processes requires parameters such as rate constants, partition coefficients or similar experimentally measurable values. The combination of these inputs, the representations of processes, and the model spatiotemporal resolution dictates the resulting model outputs. These models can have high explanatory power because of this structure and can directly relate changes in drivers to model output values in a bottom-up approach. The uncertainty in these models’ output values are the accumulation of the types of uncertainty discussed in section 6.4, and the model outputs are not necessarily the same quantity that monitoring efforts are measuring, but the two often overlap closely.

To conduct bottom-up analyses, estimates of primary anthropogenic emissions/releases of mercury are required as inputs for a variety of models in different media. While some inventories are currently available, they differ in methodology, represented time period, and release magnitudes. An updated, unified emission inventory which estimates both magnitudes of releases and their uncertainties would aid the Effectiveness Evaluation and provide more robust answers to questions of trend identification and attribution.

Statistical models can also be useful, especially in areas where the process-level understanding is insufficient to allow representation by mechanistic models, but where a cause-effect relationship between predictor variables and the quantity to be predicted can reasonably be justified with sound scientific explanations. When used together with mechanistic models, statistical models can be useful to determine if the process-level understanding is good enough. Such models require separate training and test data to avoid overfitting and careful determination of predictor variables to avoid confounding factors.
Statistical models trained on primary and ancillary monitoring data can have high predictive power within the range of the training (or input) data and can identify and control for variations in drivers which can obscure an underlying mercury-specific signal, but lack the explanatory ability of process-based models. This type of top-down approach can be useful for attribution on its own, especially when used in a manner that accounts for collinearity between predictor variables and predictors which may not fully capture the underlying mechanisms (e.g. Bayesian networks). Top-down approaches can also be combined with bottom-up approaches to infer a best estimate which uses both observed quantities and prior estimates.

Multiple complementary models that can be used to answer the same questions should be employed together wherever possible. This can be accomplished using a Bayesian approach, whereby bottom-up analysis (representing process-level knowledge) provides a prior estimate independent from the monitoring data itself, and top-down analysis provides the likelihood of those same estimates given the monitoring observations (representing the weight of evidence). By incorporating the quantified uncertainties of each model into this approach, a more robust estimate can be obtained. In many cases this can result in lower overall uncertainty in the quantitative answer to a given question by combining the higher predictive power of top-down approaches with the higher explanatory power of bottom-up approaches. This can be viewed as a way for statistical models to “calibrate” mechanistic models based on the observational findings from monitoring, in a way that is specific to a given question and uses all available information. This approach can be used for a single medium with multiple applicable models and/or to combine models across media.

**Figure 6.1.** Illustration of attribution across media for hypothetical contributions of selected drivers at a hypothetical location. The coloured bars represent the fractional contributions of different drivers to observed mercury trends/variability in each medium. The drivers of variability/change in a given medium can in turn be attributable to drivers in other media (C. Thackray, unpublished).
Model Description 6.5: Mass balance models (also referred to as box models, compartment models, mass flow models)

Mass balance models represent the exchange of mercury between media, and are versatile tools which can be used on a range of spatial scales. These models use estimates of how quickly mercury is exchanged between media to self-consistently calculate mercury levels across a wide range of time scales, in a trade-off against spatially-resolved output. Inputs to these models are the releases of mercury into the model domain. Such models representing the global mercury cycle would take as inputs the total anthropogenic and geogenic releases of previously-lithospheric mercury, and represent its fate as it cycles through the atmosphere, terrestrial and ocean systems on decadal timescales and longer. In contrast, the same modelling approach could be applied to a specific location, with the inputs then being local releases of mercury as well as the transport of mercury from outside the model domain, and the model representing local mercury levels instead of global average levels. Mass flow models can be used to evaluate local effectiveness over smaller regions by representing the processes and releases particular to those regions and using the contribution of global trends as an external input. An important consideration when comparing to point observations is the spatial aggregation implied by a single or few compartments representing each entire medium in this type of model.

- Strengths: Bottom-up attribution, consistency across timescales
- Weaknesses: Reliance on wide range of input quantities
- Readiness: Easily implemented where inputs are available

Model Description 6.6: Watershed Hg models

Watershed models combine mechanistic and empirical models that each capture the dynamics of a particular component of the local biogeochemistry to simulate mercury and methylmercury concentrations and fluxes (Golden and Knightes, 2011; Knightes et al. 2014). This type of modelling is highly watershed-specific and relies on in-depth a priori knowledge of the watershed system of interest. The biogeochemical processes within the watershed contribute along with large-scale drivers such as thawing permafrost and land-use change to dictate the mercury response. Since understanding of the full collection of processes is incomplete and the local variability of the biogeochemical conditions are large, a range of ancillary parameters are therefore needed to enable statistical analysis of source-receptor relationships. This type of location-specific modelling is particularly important for sensitive environments such as the Arctic and for contaminated sites.

- Strengths: Characterize complex interactions of important processes
- Weaknesses: Intensive implementation, large uncertainty
- Readiness: Possible research implementation at intensive monitoring sites

Model Description 6.7: Food web and bioaccumulation models

Food web models represent the uptake of methylmercury to biota and the resulting bioaccumulation in freshwater and marine food webs. Inputs to these models are water concentrations of mercury and other chemical variables (e.g. DOC, TSS, pH), and parameters include water temperatures and bioenergetics parameters (and in some cases food web structures). Models either represent specific food webs and therefore simulate concentrations for species directly or simulate concentrations by
trophic level. In both cases, measured tissue concentrations and trophic position are key for evaluating these models. Important comparability considerations include the age and size of sampled individuals and movement outside the represented domain (e.g., migration). The relevant timescales for food webs are years to a decade. These models can take local observations of marine or freshwater mercury levels and trends and translate them to fish concentrations to inform local exposure modelling.

- **Strengths:** Bottom-up attribution, specificity
- **Weaknesses:** Some parameters difficult to obtain (e.g., food web structure, water biogeochemistry), challenging to extend site-specific models to larger areas
- **Readiness:** Multiple site-specific models available. Further study is necessary to determine how current food-web model would be best used in the Effectiveness Evaluation processes

### 6.2.5. Estimation of exposure and adverse impacts

Possible sources and routes of human mercury exposure can be identified by regression-based approaches that can be used to relate mercury biomarkers to survey data. Survey data can give estimates of dietary intake, occupational practices, and other potential influences on exposure. Using human monitoring and survey data (Tier 2 and 3) in combination allows a top-down identification of exposure pathways for specific subpopulations. Bottom-up estimates of exposure can also be possible for certain subpopulations, using local air and/or biota monitoring data where occupational and diet information is available.

Risk assessment techniques can be used to estimate risk for populations potentially affected by variable exposure levels. Through probabilistic relationships between human biomarker mercury and adverse health effects, subpopulation burden of disease can be estimated, and extended to economic costs. With monitoring repeated on the same subpopulations over time, the changes in these expected health effects and their costs can be quantified and related to exposure pathways.

### Model Description 6.8: Exposure and human health risk assessment models

This category encompasses a collection of models representing human exposure to mercury and the resulting health risks. Exposure models represent the intake of mercury by humans, and require mercury concentrations in diet items (e.g., freshwater fish, seafood, marine mammals, rice), in occupational practices (e.g., ASGM, dentistry), in certain products (e.g., skin-lightening creams, waste products), and the environment (e.g., soil). To mechanistically link exposure to mercury biomarker concentrations (i.e., levels in blood, urine, hair, and/or cord blood) in a given population (e.g., for a “bottom-up" analysis), toxicokinetic parameters describing human mercury metabolism can be used, with important uncertainties arising from differences in methylmercury uptake and elimination across individuals (Stern et al. 2005). Regression-based models (including GAM/GLM) are also commonly used to relate human biomarker concentrations to exposure pathways in populations and subpopulations and can be used for "top-down" attributive analysis. Human biomarkers in populations may react to changes in mercury exposure in the timescale of days to months, with documented examples related to fish consumption advisories, amalgam removal, and occupational practices. Health impacts are often important for specific sub-populations, for example people who rely on local fish and marine mammals as a dominant protein source, and people with occupational exposures such as artisanal and small-scale gold miners. Health impacts are commonly modelled using statistical relationships, with acute and chronic responses to inorganic mercury exposure and
longer-term impacts of dietary methylmercury exposure. These models can be applied at local/population scales to estimate effectiveness of changes in global mercury releases on limiting exposure and health impact, using observed biota mercury levels, or those from food web models, as inputs. Other site-specific applications are acute and chronic occupational exposures, such as at artisanal and small-scale gold mines.

- Strengths: Designed to interact with monitoring quantities, well-defined procedure
- Weaknesses: Potential recall bias in survey data needed to simulate exposure, inter-individual and -population variation in toxicokinetic parameters for mechanistic models
- Readiness: Well-established methodology

6.2.6. Quantification of key environmental processes

Top-down approaches using air monitoring and associated ancillary data (Tier 2 and 3) can estimate the contributions of specific environmental processes to observed mercury variability. Wide-ranging comparable ancillary data including land-use, air quality, and dry deposition parameters (Tier 2) and isotope measurements (Tier 3) can allow identification of their influence on observed mercury concentrations and wet deposition.

Large scale intercomparisons of monitoring measurements with bottom-up model output can also help identify key processes. Where monitoring shows inconsistency with mechanistic models, it indicates an area to better identify and quantify the important processes and their effects. Where the contribution of sources and sinks to levels of mercury is explicitly represented by mechanistic models, the observed levels, patterns, and trends can be used to infer changes in individual drivers.

In the atmosphere, oceans, and terrestrial system, observed spatial patterns of mercury and how they relate to environmental drivers can inform how modelling can best represent the physical and chemical processes that determine the transport and fate of mercury. Better representation of these processes will increase the applicable scope of the modelling and contribute to an iterative process where future modelling will better answer the Effectiveness Evaluation questions of interest.

6.3. Role of coupled-media modelling and analysis

Models or modelling frameworks that simulate multiple media and the flows of mercury between them in an internally consistent fashion are especially useful in light of the connections between media across a range of space and time scales. Each model discussed in this chapter represents the processes important to a specific medium, and these media are interconnected in a variety of ways. Some of these connections are effectively one-way, with one medium affecting another but not vice versa. In these situations, models can be chained together by using the output of a model for one medium as an input to a model for another. When models are chained in this way, longer simulations may be required to address the different retention times of different media.

On the other hand, some of the connections between media are effectively two-way, with both media affecting each other, possibly on different time scales. In these cases, coupled-media models which represent processes in both media in an internally consistent fashion are important for accurately attributing observed levels and trends to their drivers. The internal consistency can reduce uncertainty in situations where fewer of the possibilities for individual media are consistent across multiple media at
once. The representation of coupling across multiple timescales means that these models can be more applicable for longer-term trends influenced by legacy mercury. The two-way coupling of existing single-medium models can be technically challenging, depending on the model specifics and time scales involved.

While the response of the atmosphere to changes in air emissions is relatively fast, on the order of months to years, the response in other media can be slower and lag behind those changes. Moreover, the responses of the terrestrial and ocean systems feedback on the atmosphere, causing atmospheric trends to contain a signal contributed on these longer timescales. On the global scale, a decrease in anthropogenic emissions of mercury to air results in a fast atmospheric response proportional to the change in the total flux of mercury to air, which includes significant contributions from land and ocean legacy emissions. The immediate response of the atmosphere is thereby dampened by the slower-equilibrating media. For example, declining atmospheric concentrations result in declining deposition to both land and oceans. This declining deposition leads to a decline in mercury levels in those media on longer time scales, which itself leads to less mercury evaded to the atmosphere and further declines in atmospheric mercury. Models which provide a coupled atmosphere-ocean-terrestrial simulation can be used for modelling trends of atmospheric, terrestrial, and ocean mercury concentrations simultaneously. This would be particularly useful for identification and attribution of trends influenced by legacy mercury.

Coupled-media models can help us to understand the implications of the trends we observe in air or other media for the eventual impacts on ecosystems and humans, which will be manifested over time. Observed decreases in air concentrations and deposition will likely contribute to decreased human exposure in the future. Even though we cannot yet observe those benefits, coupled-media models can be used to estimate them.

6.4. Model uncertainty

Some types of model uncertainty can be quantified, and represented as a distribution of the probability of specific values. This allows models to not only estimate quantities of interest, but also to provide a measure of confidence in those estimates. This is important for basing decisions and evaluations on model results, and for identifying which policy-relevant questions have clear answers and which require further monitoring/analysis to answer with a given degree of confidence. Combining multiple modelling approaches with well-quantified uncertainties can reduce overall uncertainty by identifying areas of agreement.

The common structure of mechanistic models produces model output with three important categories of uncertainty that should be considered in model evaluation and interpretation of “bottom-up” estimates:

a) uncertainty in the output which follows from the fact that the inputs used by the model are themselves uncertain (e.g. inventories of emissions/releases). This uncertainty can be estimated by testing a model using a range of available estimates of the inputs.

b) uncertainty in the output which follows from the fact that the physical/chemical parameters used to represent different processes are uncertain (e.g. reaction rate coefficients and partitioning coefficients). This uncertainty can be estimated by testing a model using a range of parameter values within their uncertainty bounds.
c) structural uncertainty due to the fact that there are processes and levels of mechanistic complexity that are not represented by the model due to incomplete knowledge about the drivers of the behaviour of modelled quantities. This type of uncertainty can be difficult to quantify because it potentially depends on unknown missing processes, but can be qualitatively assessed by experts with knowledge of the processes represented by the model.

Uncertainty for generalized linear and additive models can be estimated based on the standard error of predictions of observations in a cross-validation dataset. Confidence intervals can also be calculated using the posterior distribution of the model parameters.

In addition to model uncertainty, the comparison of modelled and observed quantities requires consideration of uncertainty and variability in the observational data, and the uncertainty due to the comparison itself. The latter can be introduced through mismatches in the precise nature of the quantities compared, especially via spatial and temporal mismatches. The comparison of gridded model output with point observations introduces such uncertainty, because within the area of a model grid cell some unresolved variability is to be expected. The magnitude of this uncertainty can be estimated if point observations at multiple locations within a single model cell are available, or by downscaling larger-scale variability in model output. Mismatches between model temporal resolution and observational sampling frequency and averaging time similarly need to be considered. All model outputs and observational data carry uncertainty, and the quantification of that uncertainty allows decision-making to be based not just on a given result, but also the degree of confidence in that result.

6.5. Model evaluation

Where possible, multiple applicable models can be used together for increased robustness rather than selecting a single model for a particular question. The evaluation of a model for use requires the determination of under what conditions and for what quantities/questions that model is applicable. The quantities for which the model is used should be directly calculated by the model, and the model should generate results that are consistent with directly comparable monitoring/observational data. In comparisons of models to monitoring mercury levels and trends, model-monitoring equality should not be the goal, but rather model-monitoring consistency. For quantities of interest, model and measured values are consistent if they are not statistically distinguishable from each other when accounting for the uncertainties in the model, the measurement, and the manipulation of each for the purpose of comparison. In order to draw conclusions from the applied model, the uncertainty in its results must be smaller than the magnitude of the result itself.

<table>
<thead>
<tr>
<th>Model Evaluation Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicability of a model estimating a given quantity required to answer a question of interest should be determined by:</td>
</tr>
<tr>
<td>1. Whether that quantity is estimated by the model directly, using relationships to input variables soundly based on available knowledge</td>
</tr>
<tr>
<td>2. Whether the modelled quantity is consistent with available comparable monitoring results</td>
</tr>
<tr>
<td>3. Whether the uncertainty in the modelled quantity is well-quantified and small enough to draw the conclusions necessary to answer the question of interest and/or provide a degree of confidence in that answer.</td>
</tr>
</tbody>
</table>
### 6.6. Summary of information from modelling

Table 6.3 summarizes what output models can produce to support the Effectiveness Evaluation. Model data formatted and managed for interoperability/harmonization with both monitoring data and other models, following the FAIR criteria described in chapter 2, would greatly facilitate both single-medium and multi-media analyses. A common, self-describing and open data format should be used for gridded model output so that data users can rely on a single set of free and open software tools for all shared model data. Shared model output should include quantities for comparison to monitoring as well as quantities that are common inputs to other models, such as fluxes across media boundaries, as well as metadata containing relevant information about the output and how it is generated.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Primary Output</th>
<th>Evaluation Output</th>
<th>Metadata</th>
<th>Location Data</th>
<th>Output For Other Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmosphere</td>
<td>Air concentrations and temporal trends</td>
<td>Atmospheric concentrations, trends; wet deposition rates, trends; dry deposition to foliage/soils/snowpack</td>
<td>Input sources; meteorological inputs; chemistry represented; boundary assumptions</td>
<td>Latitudes, longitudes, altitudes</td>
<td>Gross dry deposition of elemental and oxidized mercury to terrestrial and ocean locations; elemental mercury concentrations; source attribution quantification for outputs</td>
</tr>
<tr>
<td>Ocean</td>
<td>Seawater concentrations and temporal trends</td>
<td>Seawater mercury concentrations, temporal trends</td>
<td>Input sources; circulation source; processes represented;</td>
<td>Latitudes, longitudes, depths</td>
<td>Gross evasion fluxes to air or seawater surface elemental mercury concentration; seawater methylmercury concentrations</td>
</tr>
<tr>
<td>Terrestrial</td>
<td>Soil/vegetation mercury levels and temporal trends</td>
<td>Soil/vegetation mercury reservoirs, trends; soil-air fluxes, temporal trends</td>
<td>Input sources; meteorological/climate inputs; processes represented;</td>
<td>Latitudes, longitudes</td>
<td>Gross evasion fluxes to air</td>
</tr>
<tr>
<td>Watershed</td>
<td>Water mercury and methylmercury concentrations</td>
<td>Freshwater mercury levels, temporal trends</td>
<td>Input sources; biogeochemical conditions; land-use</td>
<td>geolocation or represented watershed</td>
<td>Water mercury and methylmercury concentrations</td>
</tr>
</tbody>
</table>

**Table 6.3.** Information from modelling data. For each model type, the primary model output is listed, along with the output appropriate for evaluation, metadata to accompany the model output, data for identifying model output locations, and model output to be collected for use by other types of models.
### 6.7. Conclusions

By analysing monitoring data, temporal and spatial trends in the levels of mercury in specific environmental media or human matrices can be derived with confidence intervals. These trends provide a first-level indication of whether the Convention may be contributing to protecting human health and the environment from the adverse effects of mercury. Analyses of the monitoring data collected in each medium separately will be informative, and these monitoring data can also be used in an integrated manner, where combining multiple complementary analysis approaches to answer the same question will improve robustness and increase the scientific weight of evidence. This is particularly important when models are used for policy evaluation applications and uncertainties need to be quantified and minimized. Both for model evaluation and for analyses, model output uncertainties should be quantified. Any generated estimate should provide a detailed discussion/presentation of the associated uncertainty and factors that have determined this uncertainty.

In many cases, attribution of observed trends to specific drivers can be performed through the use of models which resolve the intervening processes, supplemented by empirical statistical approaches. Cross-media analysis involving both mechanistic and statistical modelling in all relevant media is
important in order to fully evaluate effectiveness of the Convention. This evaluation requires separating the impacts attributable to the Convention from changes that occur due to other factors, and while monitoring data shows the impact of all of these factors, modelling can help attribute the changes to the different drivers. As more monitoring data and analysis tools become available, more detailed analysis can be performed.

To estimate background and impacted levels of mercury, simple analyses can be conducted on monitoring data at sites chosen for this purpose. Temporal trends can be identified at these and other locations once a long enough time record has been collected. This trend analysis should account for variability and uncertainty to obtain trend magnitudes, confidence intervals for the trends, and measures of the trends’ statistical significance.

To characterize spatial patterns, several atmospheric chemical transport models can be used, supplemented with statistical models where beneficial to quantify representativeness of observed levels and trends in air, and to extrapolate ambient air concentrations and wet deposition to areas with sparse monitoring data. Spatially resolved models in air and other media can be used to interpolate levels and trends of mercury while accounting for the drivers of spatial and temporal differences.

Bottom-up analyses can be performed with atmospheric models for source attribution, and top-down analyses with GLM/GAM for air and biota attribution where sufficient ancillary data is available. Top-down analysis of changes in exposure pathways can also be performed to attribute changes in human biomarkers to measures influenced by the Convention. At intensive monitoring sites, combined top-down and bottom-up attribution analyses can be performed for air, biota and human biomarkers. To quantify legacy impacts, coupled-media approaches should be used where possible.

Exposure can be estimated based on specific sources and exposure attribution information can be used to estimate marginal health impacts/costs of individual drivers. Trends in risk associated with trends in exposure and/or biomarker benchmark values can be estimated where the appropriate information is available.

The quantification of key environmental processes can improve our understanding of cause-effect relationships. Top-down analysis can be used to identify key environmental drivers, and large-scale measurement/model intercomparisons can be performed to identify key processes. Improved understanding can lead to a beneficial iterative approach: using the available information to improve the application of models can decrease the uncertainty for further and future analyses and evaluations.

A.1. Introduction

To support Parties and organizations who may wish to develop new monitoring programs, or improve existing ones, with a view to contributing to the Effectiveness Evaluation, this document identifies a tiered approach for monitoring each of the three media (air, biota, humans):\textsuperscript{16}

Tier 1 is intended to provide guidance on mercury monitoring under a limited set parameters for circumstances where available resources are not sufficient to implement the actions in Tier 2. Following guidance by the COP,\textsuperscript{17} the methods in Tier 1 are cost effective, practical, feasible, and sustainable. The Tier 1 methods are intended to provide information that are useful in identifying and characterizing gaps and needs of national, regional, or local interest and to provide information that is useful to the collective effort for the Effectiveness Evaluation. While the implementation of Tier 1 actions may not fully address the questions in Table 2.1, it will contribute essential information and create a foundation for Tier 2 monitoring.

Tier 2 is intended to build upon Tier 1 methods to provide information that will address the questions identified in Table 2.1, and to create a basis for assessing source attribution at the local, national, and global scales (Figure 2.2). The methods and approaches in this tier may be more expensive or complex than those under Tier 1. The more comparable data from Tier 2 becomes available, the more robust the Effectiveness Evaluation will be.

Tier 3 identifies research methods and approaches that may play a vital role in supporting the Tier 1 and Tier 2 programs and the Effectiveness Evaluation, primarily by improving our understanding of key processes that link sources to environmental concentrations and exposures. Because Tier 3 focuses on processes, the results would likely yield insights that are broadly applicable and that should be taken into consideration in the Effectiveness Evaluation when available.

An overview of a proposed tiered approach for each matrix (air, biota and Human) is shown below.

\textsuperscript{16} It is noted that the Convention does not impose any obligation upon Parties to conduct monitoring. As such, the tiered approach and any other activities or recommendations contained in this guidance are voluntary and presented with the sole purpose of supporting Parties who may wish to develop new monitoring programs, or improve existing ones, with a view to contributing to the Effectiveness Evaluation.

\textsuperscript{17} Decision MC-2/10 pursuant to the terms of reference to Ad-hoc Technical Expert Group on Effectiveness Evaluation.
### A.2. Atmospheric mercury monitoring

<table>
<thead>
<tr>
<th>Hg Measurement</th>
<th>Ancillary Measurements</th>
<th>Location/Spatial Distribution</th>
<th>Frequency</th>
<th>Contribution to information categories(^{18})</th>
<th>Modelling/Analysis(^{19})</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TIER 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Gaseous Mercury or Gaseous Elemental Mercury (a range of methods may be used depending on objectives, resources, and other constraints): - continuous mercury analysers (recommended where possible) - manual trap methods - passive samplers</td>
<td>Precipitation, meteorological data (where available, may be from nearby sites).</td>
<td>Sites should be selected in a mix of locations that include a) remote, background, b) regionally representative, and c) source impacted locations. Siting strategies may differ if the methods deployed are only active, only passive, or a mix of active and passive. Deploying a mix of active and passive samplers may maximize the amount of information collected given resource, infrastructure, or personnel constraints. Where possible, measurements should be collocated with other types of air quality and mercury measurements.</td>
<td>Varies by method.</td>
<td>(1) Baselines and Spatial Patterns; (2) Temporal Trends; (5) Estimating Exposure and Adverse Effects.</td>
<td>(1a) TGM/GEM concentrations from comparable measurements (present aggregate results per site type); supplemented by modelled TGM/GEM concentrations; (1b) Wet deposition concentrations from comparable measurements (present aggregate results per site type); supplemented by modelled wet deposition for areas without measurements; (2) Mann-Kendall, Sen's slope; (4a) &quot;Bottom-up&quot; CTMs for TGM/GEM. (4b) &quot;Bottom-up&quot; CTMs for wet deposition. Measurement / model intercomparisons for ambient mercury concentrations and wet deposition.</td>
</tr>
</tbody>
</table>

\(^{18}\) See chapter 2.

\(^{19}\) See chapter 6.
<table>
<thead>
<tr>
<th>TIER 2</th>
<th>TIER 3</th>
</tr>
</thead>
</table>
| Speciated Reactive Mercury:  
- high resolution measurements of PBM2.5, GOM using existing network SOPs;  
- cation exchange membranes. 

Dry deposition of mercury:  
- Total Hg and MeHg litterfall and throughfall (select forest ecosystems). 

Emission inventories, atmospheric deposition of sulfate, land cover, land use, leaf area index, meteorology, air quality tracers (including SO₂, CO₂, CO, PM2.5).  

Expect a few sites in each world region, surrounded by a cluster of Tier 1 sites. Sites should be a mix of a) remote, background; b) regionally representative; and c) source impacted locations and collocated with other network sites with more robust infrastructure. 

Varies by method; high temporal resolution for speciation. 

(1) Baselines and Spatial Patterns;  
(2) Temporal Trends;  
(3) Key Environmental Processes;  
(4) Source Attribution;  
(5) Estimating Exposure and Adverse Effects. 

| Mercury Isotopes:  
- e.g. multi-collector inductively coupled plasma;  
- mass spectrometry (MC-ICP-MS) 

Additional speciation methods  
Applications of Tier 1 and Tier 2 methods in intensive research contexts to support process understanding 

Same as above. 

Expected to be opportunistic siting, collocated at long-term monitoring and research sites.  

Aircraft campaigns, ocean surveys, flux towers, etc. 

Data is collected through research programs and will be provided when available for the purposes of supporting the Effectiveness Evaluation. 

(2) Temporal Trends;  
(3) Key Environmental Processes;  
(4) Source Attribution;  
(5) Estimating Exposure and Adverse Effects. 

(4) Calibration of "Bottom-up" CTMs input parameters with "top-down" approaches; Including isotopic fingerprints, where applicable. Measurement-model fusion to estimate total mercury deposition; dry deposition modelling (local).
# A.3. Biota monitoring

<table>
<thead>
<tr>
<th>Hg Measurement</th>
<th>Ancillary Measurements</th>
<th>Location/Spatial Distribution</th>
<th>Frequency</th>
<th>Contribution to information categories</th>
<th>Modelling/Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIER 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue from bird and fish can be used to monitor total mercury levels in biota. Species to be selected for monitoring should have, where possible, a relatively consistent diet (and thus a narrow trophic range) that can be observed consistently over time at a given location. Trophic level 4 species are used in a number of existing programs and are a reasonable starting point.</td>
<td>Spatial Coordinates (Latitude/Longitude), Species Name, Body Length &amp; Mass, Age, Sex and Maturity Stage; catchment description (size of lake, elevation, landcover and use)</td>
<td>It is most important to make consistent observations at fixed locations over a long period. A mixture of background sites and locally impacted sites is recommended. With sufficient prior information, sites with well-known impact history should be chosen. Where little or no prior information exists, the possibility of using ecosystem sensitivity modelling for the chosen sites may be explored.</td>
<td>Annual measurements, with a consistent sampling season over time for each core fixed site.</td>
<td>(2) Temporal Trends; (5) Estimating Exposure and Adverse Effects.</td>
<td>(2) Time series analysis of individual observations within a generalized linear model (GLM) or framework (i.e. mixed effects model) that can account for variation associated with species and site. Note that individual sample data is most useful for analysis rather than aggregated values.</td>
</tr>
<tr>
<td>TIER 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Tier 2, a consistent taxon would be sampled in different sites over time. While it is important to sample as consistent a taxon as possible across locations, if that is not possible, sampling several taxa in the multiple sites would help in accounting for species effects.</td>
<td>Additional ancillary measurements may include: In biota: carbon (δ13C) &amp; nitrogen (δ15N) stable isotopes; In water: DOM/DOC/TOC, TSS, salinity, DO, (pH). N and P, phytopigments (e.g. chlorophyll-a); In sediment: THg In Air: GEM, wet deposition, and meteorological data.</td>
<td>Sites added in this tier would be sampled to cover a wider range of landscapes and geochemical characteristics. The additional sites may be selected, for example, according to the type of habitat type and then either rotated or randomly sampled within each habitat category. If the data sets from</td>
<td>Yearly monitoring rotating across sites added at Tier 2 (in such a manner that each particular site would only be monitored every few years).</td>
<td>(1) Baselines and Spatial Patterns; (2) Temporal Trends; (3) Key Environmental Processes; (4) Source Attribution; (4a) &quot;Top-down&quot; approach: Use generalized additive model (GAM) or Bayesian models to estimate the relative contribution of different influxes, sources and large-scale drivers with the help of the ancillary measurements (assumed to be affected by them). Local and regional</td>
<td></td>
</tr>
</tbody>
</table>

---

20 See chapter 4.  
21 See chapter 8.
statistically. However, it is noted that monitoring novel species that have not been previously monitored elsewhere would be less informative for the Effectiveness Evaluation.

| Description of local hydrologic catchment. | additional locations are paired with those from fixed sites monitoring similar covariates over time, the combined data sets will inform each other and contribute to source attribution. If possible, air and deposition measurements should also be carried out for the same sites. | (5) Estimating Exposure and Adverse Effects. | influxes / sources / drivers can be presented in a combined form to form a global picture; (4b) "Bottom-up" approach: Use mechanistic models to distinguish between sources of deposited mercury. |

<table>
<thead>
<tr>
<th>TIER 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling as above, but consideration may be given to species in lower trophic level taxa. Species at lower trophic levels may provide useful information to attribution of changes as they are more likely to respond more quickly to changes in Hg exposure and show changes earlier.</td>
</tr>
<tr>
<td>Mercury (δ202Hg) stable isotopes in biota and suspected source-matrixes of interest; other chemical tracers related to known drivers (i.e. changes in CO2 levels and water temperature in oceans due to climate change, co-tracers from ASGM activity, etc.). Information on diet (e.g. fatty acids), stable isotopes of lower foodweb organisms (or compound specific stable isotopes of amino acids in fish), data on food web structure, as well as associated land cover data.</td>
</tr>
<tr>
<td>Intensively monitor selected areas (e.g. catchments), with a primary site (supersite) for collocated measurements and secondary (or satellite) sites to capture variability across the catchment. Catchments selected for this strategy may be either background locations (mostly influenced by long range transport) or locally impacted locations (that are likely to see changes due to mitigation efforts).</td>
</tr>
<tr>
<td>Sampling may be more frequent than annual.</td>
</tr>
<tr>
<td>(1) Baselines and Spatial Patterns; (2) Temporal Trends; (3) Key Environmental Processes; (4) Source Attribution; (5) Estimating Exposure and Adverse Effects.</td>
</tr>
<tr>
<td>(4) Calibrate &quot;Bottom-up&quot; CTMs input parameters with &quot;top-down&quot; approaches, including isotopic fingerprints, where applicable.</td>
</tr>
</tbody>
</table>
## A.4. Human biomonitoring

<table>
<thead>
<tr>
<th>Hg Measurement</th>
<th>Ancillary Measurements</th>
<th>Location/Spatial Distribution</th>
<th>Frequency</th>
<th>Contribution to information categories$^{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood, urine, or hair THg depending on sampled population.</td>
<td>WHO Survey or HBM4EU Instruments. Relevant survey information (e.g., dietary, occupational, sociodemographic), where possible.</td>
<td>Vulnerable sub-populations should be identified based on exposure or risk that is most critical for them (i.e. dietary exposures, occupational groups, or high risk lifestage (e.g. pregnant women)).</td>
<td>Every 2-5 years for the same population, with monitoring activities staggered for different populations in different years.</td>
<td>(1) Baselines and Spatial Patterns; (2) Temporal Trends; (5) Estimating Exposure and Adverse Effects.</td>
</tr>
</tbody>
</table>

**TIER 2**

| Blood/cord blood, urine, and/or hair THg depending on sampled population and survey. Methyl mercury and isotopes may also be considered. | WHO Survey or HBM4EU Instruments, or incorporation of Hg sampling into other health surveys or cohort studies. Relevant survey information (e.g., dietary, occupational, sociodemographic), and where possible coordinated measures in air and/or biota. | Two strategies: 1) Perform more in-depth analysis of sub-populations with high-exposure or classified as a vulnerable lifestage; 2) Incorporation of Hg sampling into other, in-depth health surveys or cohort studies. Same as above. | (1) Baselines and Spatial Patterns; (2) Temporal Trends; (4) Source Attribution; (5) Estimating Exposure and Adverse Effects. |

**TIER 3**

| same as above for Tier 2 | WHO Survey or HBM4EU Instruments or National/Regional population survey instruments. Relevant survey information (e.g., dietary, occupational, sociodemographic), and where possible coordinated measures in air and/or biota. | Two strategies: 1) National population survey (ideally leveraging other surveys/samples, and inclusion of vulnerable sub-groups); 2) Sampling of sub-populations with coordinated air and biota sampling. Same as above. | (1) Baselines and Spatial Patterns; (2) Temporal Trends; (3) Key Environmental Processes; (4) Source Attribution; (5) Estimating Exposure and Adverse Effects. |

---

$^{22}$ See chapter 4.
References [to be finalized]


AMAP. 2018. AMAP Assessment 2018: Biological effects of contaminants on Arctic Wildlife and Fish. Arctic Monitoring and Assessment Programme (AMAP), Tromso, Norway. 84pp.


Angot, H., Barret, M., Magand, O., Ramonet, M., Dommergue, A. A 2-year record of atmospheric mercury species at a background Southern Hemisphere station on Amsterdam Island. Atmospheric Chemistry and Physics, 14, 11461-11473, 2014.


